

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 06:35:38 ; Search time 60 seconds  
(without alignments)  
7620.918 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gayggggagcagtgtaacga.....gctgtgtatagcagcgagca 1491

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCITUS.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                | Description        |
|------------|-------|-------------|--------|-------------------|--------------------|
| 1          | 42.4  | 2.8         | 2518   | US-09-433-699-3   | Sequence 3, Appl1  |
| 2          | 42    | 2.8         | 486    | US-08-263-413-23  | Sequence 23, Appl1 |
| 3          | 42    | 2.8         | 500    | US-08-263-413-22  | Sequence 2, Appl1  |
| 4          | 42    | 2.8         | 675    | US-07-807-043B-2  | Sequence 2, Appl1  |
| 5          | 42    | 2.8         | 675    | US-08-299-849B-2  | Sequence 2, Appl1  |
| 6          | 42    | 2.8         | 675    | US-08-142-368A-2  | Sequence 2, Appl1  |
| 7          | 42    | 2.8         | 675    | US-08-967-727-2   | Sequence 2, Appl1  |
| 8          | 42    | 2.8         | 675    | US-08-037-230D-2  | Sequence 2, Appl1  |
| 9          | 42    | 2.8         | 1365   | US-07-807-043B-4  | Sequence 4, Appl1  |
| 10         | 42    | 2.8         | 1365   | US-08-299-849B-4  | Sequence 4, Appl1  |
| 11         | 42    | 2.8         | 1365   | US-08-142-368A-4  | Sequence 4, Appl1  |
| 12         | 42    | 2.8         | 1365   | US-08-967-727-4   | Sequence 4, Appl1  |
| 13         | 42    | 2.8         | 4698   | US-07-807-043B-5  | Sequence 5, Appl1  |
| 14         | 42    | 2.8         | 4698   | US-08-299-849B-5  | Sequence 5, Appl1  |
| 15         | 42    | 2.8         | 4698   | US-08-142-368A-5  | Sequence 5, Appl1  |
| 16         | 42    | 2.8         | 4698   | US-08-967-727-5   | Sequence 5, Appl1  |
| 17         | 42    | 2.8         | 4698   | US-08-037-230D-5  | Sequence 5, Appl1  |
| 18         | 42    | 2.8         | 150    | US-07-829-461A-8  | Sequence 8, Appl1  |
| 19         | 40.6  | 2.7         | 390    | US-09-197-646-7   | Sequence 7, Appl1  |
| 20         | 40.6  | 2.7         | 3489   | US-08-728-323A-1  | Sequence 1, Appl1  |
| 21         | 40    | 2.7         | 32207  | US-09-258-568-1   | Sequence 1, Appl1  |
| 22         | 40    | 2.7         | 32207  | US-08-770-379-20  | Sequence 20, Appl1 |
| 23         | 40    | 2.7         | 32207  | US-08-757-669A-20 | Sequence 20, Appl1 |
| 24         | 40    | 2.7         | 32207  | US-09-230-371A-20 | Sequence 20, Appl1 |
| 25         | 38.6  | 2.6         | 1186   | US-08-368-236-2   | Sequence 2, Appl1  |
| 26         | 38.6  | 2.6         | 1011   | US-08-686-528A-1  | Sequence 1, Appl1  |
| 27         | 39.4  | 2.6         |        |                   |                    |

|    |      |     |      |   |                     |                     |
|----|------|-----|------|---|---------------------|---------------------|
| 28 | 39.4 | 2.6 | 1011 | 4 | US-09-456-287-1     | Sequence 1, Appl1   |
| 29 | 39.2 | 2.6 | 1189 | 5 | US-07-781-034-4     | Sequence 4, Appl1   |
| 30 | 39.2 | 2.6 | 1189 | 5 | PCT-US92-08328-4    | Sequence 4, Appl1   |
| 31 | 38.8 | 2.6 | 1011 | 4 | US-08-487-429A-10   | Sequence 10, Appl1  |
| 32 | 38.8 | 2.6 | 1011 | 5 | PCT-US96-05320A-119 | Sequence 119, Appl1 |
| 33 | 38.8 | 2.6 | 2295 | 1 | US-08-375-300-3     | Sequence 3, Appl1   |
| 34 | 38.8 | 2.6 | 2295 | 3 | US-09-177-431-3     | Sequence 3, Appl1   |
| 35 | 38.8 | 2.6 | 2295 | 5 | PCT-US95-16930-3    | Sequence 3, Appl1   |
| 36 | 38.8 | 2.6 | 2295 | 5 | PCT-US95-16930-3    | Sequence 3, Appl1   |
| 37 | 38.8 | 2.6 | 4080 | 1 | US-08-375-300-1     | Sequence 1, Appl1   |
| 38 | 38.8 | 2.6 | 4080 | 3 | PCT-US95-16930-1    | Sequence 1, Appl1   |
| 39 | 38.4 | 2.6 | 2581 | 4 | US-09-370-838-66    | Sequence 66, Appl1  |
| 40 | 38.2 | 2.6 | 455  | 1 | US-08-843-521-3     | Sequence 3, Appl1   |
| 41 | 38.2 | 2.6 | 455  | 4 | US-09-012-871-3     | Sequence 3, Appl1   |
| 42 | 38.2 | 2.6 | 2350 | 4 | US-08-843-521-1     | Sequence 1, Appl1   |
| 43 | 38.2 | 2.6 | 2350 | 4 | US-09-012-871-1     | Sequence 1, Appl1   |
| 44 | 38.2 | 2.6 | 2882 | 4 | US-08-961-527-140   | Sequence 140, Appl1 |
| 45 | 38.2 | 2.5 | 8920 | 2 | US-08-446-855A-1    | Sequence 1, Appl1   |

## ALIGNMENTS

```
RESULT 1
US-09-433-699-3
; Sequence 3, Application US/09433699B
; Patent No. 6165786
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF NUCLEOLIN EXPRESSION
; FILE REFERENCE: RTS-0109
; CURRENT APPLICATION NUMBER: US/09/433,699B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 2518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(2235)
US-09-433-699-3
```

Query Match 2.88; Score 42.4; DB 4; Length 2518;  
Best Local Similarity 57.98; Pred. No. 0.017;  
Matches 73; Conservative 1; Mismatches 52; Indels 0; Gaps 0;

```
OY 1 GATGGGAGCAGTGTACGATGGGACAGCAAGATGACCCACCATGACCCACCATGAT 60
||: || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 814 GATGAAGATGAAGAAGAGATGATGACGAGATGACGAGACGAGACGAGATGATGAA 873
||: || | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
||: || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 874 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
||: || | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 121 GAAGCA 126
||: || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 934 GAAGCA 939
```

```
RESULT 2
US-08-263-413-23
; Sequence 23, Application US/08263413
; Patent No. 5747246
; GENERAL INFORMATION:
; APPLICANT: FANREITER, CHRISTOPHE
; APPLICANT: COCHET, MADELINE
; APPLICANT: DARCHE, SYLVIE
; APPLICANT: KOURILSKY, PHILIPPE
; TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
; TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF INTEREST BY A
; TITLE OF INVENTION: METHOD OF ENZYMIC AMPLIFICATION OF DNA
```

```

1  NUMBER OF SEQUENCES: 36
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
4  ADDRESSEE: P.C.
5  STREET: 1755 Jefferson Davis Highway, Fourth Floor
6  CITY: Arlington
7  STATE: Virginia
8  COUNTRY: U.S.A.
9  ZIP: 22202
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: Patent In Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/263,413
17 FILING DATE:
18 CLASSIFICATION: 435
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 07/882,980
21 FILING DATE: 14-MAY-1992
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Ivalley, J.P.
24 REGISTRATION NUMBER: 31,451
25 REFERENCE/DOCKET NUMBER: 660-058-55X
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (703)521-4500
28 TELEFAX: (703)666-2347
29 TELETYPE: 248855 OPAT UR
30 INFORMATION FOR SEQ. ID NO: 23:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 496 base pairs
33 TYPE: nucleic acid
34 STRANDEDNESS: unknown
35 TOPOLOGY: unknown
36 MOLECULE TYPE: DNA (genomic)
37 US-08-263-413-23

```

```

Query Match      2.8%; Score 42; DB 1; Length 496;
Best Local Similarity 65.2%; Pred. No. 0.0093;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

QY      1  GAYGGGAGCAGCGTGTAAACGATGGCGACAGCAAAAGATGACACACCATGAGCAGCACCAGCAT 60
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      250 GTCGATGAGCATGAGACGCGATGAGGATGATGATGATGATGATGATGATGATGATGAT 309

QY      61  GATCCACCATGACGACCATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 92
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      310 GACGACGATGCTTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 341

RESULT 3
US-08-263-413-22
: Sequence 22, Application US/08263413
: Patent No. 5747246
: GENERAL INFORMATION:
: APPLICANT: PANNETIER, CHRISTOPHE
: APPLICANT: COCHET, MADELEINE
: APPLICANT: DARGHE, SYLVIE
: APPLICANT: KOURILSKY, PHILIPPE
: TITLE OF INVENTION: PROCESS FOR DETERMINING THE QUANTITY OF
: TITLE OF INVENTION: A DNA FRAGMENT OF INTEREST BY A METHOD OF
: TITLE OF INVENTION: METHOD OF ENZYMATIC AMPLIFICATION OF DNA
: NUMBER OF SEQUENCES: 36
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT,
: STREET: 1755 Jefferson Davis Highway, Fourth Floor
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:

```

```

1  MEDIUM TYPE: Floppy disk
2  COMPUTER: IBM PC compatible
3  OPERATING SYSTEM: PC-DOS/MS-DOS
4  SOFTWARE: Patentin Release #1.0, Version #1.25
5  CURRENT APPLICATION DATA:
6  APPLICATION NUMBER: US/08/263,413
7  FILING DATE:
8  CLASSIFICATION: 435
9  PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 07/882,980
11 FILING DATE: 14-MAY-1992
12 ATTORNEY/AGENT INFORMATION:
13 NAME: lawall, lye, j. p.
14 REGISTRATION NUMBER: 31,451
15 REFERENCE/DOCKET NUMBER: 660-058-55X
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: (703)521-4500
18 TELEFAX: (703)486-2347
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1
```

```

Query Match: 2.8%; Score 42; DB 1; Length 500;
Best Local Similarity 65.2%; Pred. No. 0.0093;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0.

QY      1  GAVGGGAGCAGCTGTACGATGGCGAACAAGATGCACCACATGACGCCACCAAGT 60
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      254 GTGATGATGAGATTAACACCATGAGCATGTAGCATGTACTCTACTGCACGACGAGGACGC 313
OY      61  GATCACCATGACGACCATGATGATGATGATGA 92
           || | |||| | ||||||| ||||| ||||| ||||| ||||| |||||
DB      314 GACGACGATGCTTTATGATGATGAGGATGA 345

RESULT 4
US-07-807-043B-2
; Sequence 2, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thlerry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807,043B
FILING DATE: 19911212
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
```



```

1 REGISTRATION NUMBER: 30,946
2 REFERENCE/DOCKET NUMBER: LUD 5253.4-US
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (212) 688-9200
5 TELEFAX: (212) 838-3884
6 INFORMATION FOR SEO ID NO: 2:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 675 base pairs
9 TYPE: nucleic acid
10 STRANDEDNESS: single
11 TOPOLOGY: linear
12 MOLECULE TYPE: genomic DNA
13
14 US-08-142-3684-2

```

[illegible]

RESULT 7  
US-08-967-727-2  
Sequence 2, Application US/08967727  
Patent No. 6025474  
GENERAL INFORMATION:  
APPLICANT: Gauglier, B atrice; Van den Eynde, Beno t;  
APPLICANT: van der Bruggen, Piere; Boon-Falleur, Thierry  
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
Zip: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,727  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/037,230  
FILING DATE: 26-MARCH-1993  
APPLICATION NUMBER: PCT/US92/04354  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/807,043  
FILING DATE: 12-DECEMBER-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/764,365  
FILING DATE: 23-SEPTEMBER-1991  
PRIOR APPLICATION DATA: 07/728,838  
FILING DATE: 9-JULY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/705,702  
FILING DATE: 23-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 6025474man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5353

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: (212) 688-9200
? TELEFAX: (212) 838-3884
? INFORMATION FOR SEQ ID NO: 2:
? SOURCE CHARACTERISTICS:
? LENGTH: 675 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: genomic DNA
? IS-Oligo-Nucleotide?: false
? IS-ORF-Nucleotide?: true

```

|         |                       |   |                  |            |                       |
|---------|-----------------------|---|------------------|------------|-----------------------|
|         | Query Match           | 2.8%  | Score 42;        | DB 3;      | Length 675;           |
|         | Best Local Similarity | 65.2%   | Pred. No. 0.011; |            |                       |
| Matches | 60;                   | Conservative  | 1;               | Mismatches | 31; Indels 0; Gaps 0; |
| QY      | 1                     | GATGGGAGACGTGTACGATGGCGACAAGAAGTGCACCCCTGTCGACGCCACCACGAT   | 60               |            |                       |
|         |                       |   |                  |            |                       |
| Dd      | 244                   | GTCGATGAGGATTAAACCATGAGGATGATGAGTTGACTCTGTCGACGACGAGGAGCAGC | 303              |            |                       |
|         |                       |   |                  |            |                       |
| QY      | 61                    | GATCACCATGACGACCATGATGATGATGATCA                            | 92               |            |                       |
|         |                       |   |                  |            |                       |
| Dd      | 304                   | GACGACGATGCCCTTCTATGATGATGATGAGTGTA                         | 335              |            |                       |

RESULT 8  
 US-08-037-230D-2  
 ? Sequence 2, Application US/08037230D  
 ? Patent No. 6235525  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;  
 ? TITLE OF INVENTION: van der Bruggen, Pierre; Boon-Pallaut, Thierry For  
 ? TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For  
 ? NUMBER OF SEQUENCES: 30  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Felfe & Lynch  
 ? STREET: 805 Third Avenue  
 ? CITY: New York City  
 ? STATE: New York  
 ? ZIP: 10022  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
 ? COMPUTER: IBM  
 ? OPERATING SYSTEM: PC-DOS  
 ? SOFTWARE: Wordperfect  
 ? CURRENT APPLICATION NUMBER: US/08/037, 230D  
 ? FILING DATE: 26-MARCH-1993  
 ? CLASSIFICATION: 435  
 ? PRIOR APPLICATION NUMBER:  
 ? APPLICATION NUMBER: PCT/US92/04354  
 ? FILING DATE: 22-MAY-1992  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 07/807,043  
 ? FILING DATE: 12-DECEMBER-1991  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 07/764,364  
 ? FILING DATE: 23-SEPTEMBER-1991  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 07/728,838  
 ? FILING DATE: 9-JULY-1991  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER: 07/705,702  
 ? FILING DATE: 23-MAY-1991  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Hanson, No. 6235525man D.  
 ? REGISTRATION NUMBER: 30,946  
 ? REFERENCE/DOCKET NUMBER: LUD 5353  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: (212) 688-9200  
 ? TELEFAX: (212) 838-3884  
 ? INFORMATION FOR SEQ ID NO: 2:



```

SEQUENCE CHARACTERISTICS:
LENGTH: 675 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-037-230D-2

Query Match
Best Local Similarity 65.2%; Pred. No. 0.011;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

2.8%; Score 42; DB 4; Length 675;
US-07-807-043B-4 Application US/07807043B
Sequence 4, Patent No. 5342774
GENERAL INFORMATION:
APPLICANT: Boon, Thierry, Van den Eynde, Beno t
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
TITLE OF INVENTION: Rejection Antigens and Uses Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807, 043B
FILING DATE: 199111212
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 09-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5342774man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 253.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 838-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: singular
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-07-807-043B-4

Query Match
Best Local Similarity 65.2%; Pred. No. 0.016;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

2.8%; Score 42; DB 1; Length 1365;
US-07-807-043B-4 Application US/07807043B
Sequence 4, Patent No. 5342774
GENERAL INFORMATION:
APPLICANT: Boon, Thierry, Van den Eynde, Beno t
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
TITLE OF INVENTION: Rejection Antigens and Uses Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/807, 043B
FILING DATE: 199111212
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 09-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5342774man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 253.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 838-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1365 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: singular
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-07-807-043B-4

```

```

Oy      1 GAAGGGACGACTGTATACGATCGGCAGAACAAATATCAACCATTACATGCCACCACCAT 60
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      706 GTCTGATGAGGATGAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 765
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      61 GATCACCATTAGACGACCATGATGATATTCATATCA 92
          || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      766 GACACGATCCCTTCTATGATGATGATGATCA 797

RESULT 10
US-08-299-849B-4
Sequence #4, Application US/08299849B
Patent No.: 5612201
GENERAL INFORMATION:
APPLICANT: De Placen, Etienne; Boon-Falleur, Thierry;
            Lech, Bernard; Salkora, Jean-Pierre; De Smet, Charles;
            Applicant Children, Patrick
TITLE OR INVENTION: Isolated Nucleic Acid Molecules Useful In
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS: 48
ADDRESSER: Felice de Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299, 849B
FILING DATE: 1-SEPTEMBER-1994
CLASSIFICATION: A45
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/037, 230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807, 043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764, 364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728, 838
APPLICATION NUMBER: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705, 702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5612201man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5355
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1365 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-299-849B-4

Query Match      2.8%; Score 42; DB 1; Length 1365;
Best Local Similarity 65.2%; Pred. No. 0.016;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;
```





```

: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Felte & Lynch
: STREET: 805 Third Avenue
: CITY: New York City
: STATE: New York
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
: COMPUTER: IBM
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,849B
: FILING DATE: 1-SEPTEMBER-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/037,230
: FILING DATE: 26-MARCH-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/04354
: FILING DATE: 22-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/807,043
: FILING DATE: 12-DECEMBER-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/764,364
: FILING DATE: 23-SEPTEMBER-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/728,838
: APPLICATION NUMBER: 9-JULY-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/705,702
: FILING DATE: 23-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 5612201man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 5355
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 838-3884
: TELEFAX: (212) 688-9200
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4698 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: US-08-299-849B-5

```

```

Query Match 2.88; Score 42; DB 1; Length 4698;
Best Local Similarity 65.28; Pred. No. 0.031;
Matches 60; Conservative 1; Mismatches 31; Indels 0; Gaps 0;
OY 1 GAYGGGAGCGAGTCTAGCATGGGACAGACCAAGATGACACCATGAGCGACACCATGAT 60
DB 706 GTGATGATGAGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 765
OY 61 GATCAGCATGAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 92
DB 766 GACGACGATGCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 797

```

Search completed: November 29, 2002, 07:42:03  
Job time : 95 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:43:55 ; Search time 19 seconds  
(without alignments)  
769.642 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497  
Sequence: 1 DGEQNDGQNKDHHDDHD.....MOGHEVESERACVIGRA 497

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2.6/ptodata/1/1aa/5A.COMB.pep.\*
- 2: /cgn2.6/ptodata/1/1aa/5B.COMB.pep.\*
- 3: /cgn2.6/ptodata/1/1aa/6A.COMB.pep.\*
- 4: /cgn2.6/ptodata/1/1aa/6B.COMB.pep.\*
- 5: /cgn2.6/ptodata/1/1aa/PCRTUS.COMB.pep.\*
- 6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                   | Description        |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1          | 7     | 1.4         | 253    | 4 US-09-149-476-576  | Sequence 576, App  |
| 2          | 7     | 1.4         | 345    | 4 US-09-194-905-12   | Sequence 12, App1  |
| 3          | 7     | 1.4         | 589    | 2 US-08-937-540-6    | Sequence 6, App1   |
| 4          | 7     | 1.4         | 590    | 4 US-09-398-395A-54  | Sequence 54, App1  |
| 5          | 7     | 1.4         | 749    | 2 US-08-937-931-8    | Sequence 8, App1   |
| 6          | 7     | 1.4         | 749    | 4 US-09-285-502-8    | Sequence 8, App1   |
| 7          | 7     | 1.4         | 749    | 4 US-09-709-126-8    | Sequence 8, App1   |
| 8          | 7     | 1.4         | 749    | 4 US-08-871-385A-8   | Sequence 103, App1 |
| 9          | 7     | 1.4         | 856    | 3 US-08-486-099-103  | Sequence 103, App  |
| 10         | 7     | 1.4         | 856    | 3 US-08-484-223B-103 | Sequence 103, App  |
| 11         | 7     | 1.4         | 856    | 3 US-08-919-597-103  | Sequence 103, App  |
| 12         | 7     | 1.4         | 856    | 3 US-08-475-668A-103 | Sequence 103, App  |
| 13         | 7     | 1.4         | 856    | 3 US-08-485-551A-103 | Sequence 103, App  |
| 14         | 7     | 1.4         | 856    | 4 US-08-471-913A-103 | Sequence 103, App  |
| 15         | 7     | 1.4         | 856    | 4 US-08-485-264A-103 | Sequence 103, App  |
| 16         | 7     | 1.4         | 856    | 4 US-08-474-349A-103 | Sequence 103, App  |
| 17         | 7     | 1.4         | 857    | 1 US-08-220-151-10   | Sequence 10, App1  |
| 18         | 7     | 1.4         | 857    | 1 US-08-413-118-10   | Sequence 10, App1  |
| 19         | 7     | 1.4         | 857    | 3 US-08-804-439A-18  | Sequence 18, App1  |
| 20         | 7     | 1.4         | 857    | 3 US-08-360-107A-113 | Sequence 113, App  |
| 21         | 7     | 1.4         | 857    | 3 US-08-473-446-10   | Sequence 10, App1  |
| 22         | 7     | 1.4         | 857    | 3 US-08-720-329-18   | Sequence 11, App1  |
| 23         | 6     | 1.2         | 9      | 4 US-09-293-322C-11  | Sequence 11, App1  |
| 24         | 6     | 1.2         | 16     | 4 US-08-602-999A-240 | Sequence 240, App  |
| 25         | 6     | 1.2         | 16     | 4 US-09-500-124-240  | Sequence 15, App1  |
| 26         | 6     | 1.2         | 27     | 3 US-08-749-816-15   | Sequence 15, App1  |
| 27         | 6     | 1.2         | 37     | 3 US-08-651-136C-64  | Sequence 64, App1  |

#### ALIGNMENTS

|    |   |     |     |   |                     |                    |
|----|---|-----|-----|---|---------------------|--------------------|
| 28 | 6 | 1.2 | 37  | 4 | US-09-229-911A-64   | Sequence 64, App1  |
| 29 | 6 | 1.2 | 59  | 4 | US-09-384-302A-16   | Sequence 16, App1  |
| 30 | 6 | 1.2 | 62  | 4 | US-09-134-001C-4865 | Sequence 4865, App |
| 31 | 6 | 1.2 | 63  | 4 | US-08-469-260A-431  | Sequence 431, App  |
| 32 | 6 | 1.2 | 72  | 4 | US-08-858-207A-443  | Sequence 443, App  |
| 33 | 6 | 1.2 | 85  | 2 | US-08-480-229C-7    | Sequence 7, App1   |
| 34 | 6 | 1.2 | 85  | 2 | US-08-659-235C-7    | Sequence 7, App1   |
| 35 | 6 | 1.2 | 90  | 4 | US-09-384-302A-17   | Sequence 17, App1  |
| 36 | 6 | 1.2 | 93  | 3 | US-09-208-804-3     | Sequence 3, App1   |
| 37 | 6 | 1.2 | 93  | 3 | US-08-801-743-3     | Sequence 91, App1  |
| 38 | 6 | 1.2 | 102 | 4 | US-09-199-637A-91   | Sequence 91, App1  |
| 39 | 6 | 1.2 | 104 | 1 | US-08-111-939A-26   | Sequence 26, App1  |
| 40 | 6 | 1.2 | 106 | 4 | US-09-199-637A-26   | Sequence 59, App1  |
| 41 | 6 | 1.2 | 108 | 2 | US-08-162-402B-27   | Sequence 27, App1  |
| 42 | 6 | 1.2 | 127 | 4 | US-09-160-246-16    | Sequence 16, App1  |
| 43 | 6 | 1.2 | 134 | 4 | US-09-201-226-2     | Sequence 2, App1   |
| 44 | 6 | 1.2 | 134 | 4 | US-09-201-227A-16   | Sequence 16, App1  |
| 45 | 6 | 1.2 | 135 | 4 | US-09-134-001C-4905 | Sequence 4905, App |

RESULT 1  
US-09-149-476-576  
; Sequence 576, Application US/0919476  
; Patent No. 6420526  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 186 Human Secreted proteins  
; FILE REFERENCE: P2002p1  
; CURRENT APPLICATION NUMBER: US/09/149,476  
; CURRENT FILING DATE: 1998-09-08  
; EARLIER APPLICATION NUMBER: PCT/US98/04493  
; EARLIER FILING DATE: 1998-03-06  
; EARLIER APPLICATION NUMBER: 60/040,162  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,333  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/038,621  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,626  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,334  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,336  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/040,163  
; EARLIER FILING DATE: 1997-03-07  
; EARLIER APPLICATION NUMBER: 60/047,600  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,615  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,597  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,502  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,633  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,583  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,617  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,618  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,503  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,592  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,581  
; EARLIER FILING DATE: 1997-05-23  
; EARLIER APPLICATION NUMBER: 60/047,584  
; EARLIER FILING DATE: 1997-05-23



```

; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match          1.4%; Score 7; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 HGYNES 307
Db 169 HGYNES 175

RESULT 2
US-09-194-905-12
; Sequence 12, Application US/09194905
; Patent No. 6306627
; GENERAL INFORMATION:
; APPLICANT: DECKER, Heinrich
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS
; TITLE OF INVENTION: GLA.O AND THEIR USE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,905
; FILING DATE: 29-JUL-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP97/02826
; FILING DATE: 30-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19622783.6
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 026083/0193
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5399
; TELEFAX: (202) 672-5300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-194-905-12

Query Match          1.4%; Score 7; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 383 QTPGST 389
Db 93 QTPGST 99

; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match          1.4%; Score 7; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 HGYNES 307
Db 169 HGYNES 175

RESULT 3
US-08-937-540-6
; Sequence 6, Application US/08937540
; Patent No. 5891697
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Wise, Mitchell L
; APPLICANT: Savage, Thomas J
; APPLICANT: Kavalita, Eva J
; TITLE OF INVENTION: Monoterpene Synthases from Common Sage
; TITLE OF INVENTION: (Salvia officinalis)
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON & KINDNESS
; STREET: 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,540
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: WSUR111254
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206 695 1718
; TELEFAX: 206 224 0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-540-6

Query Match          1.4%; Score 7; DB 2; Length 589;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 397 GPNVSD 403
Db 156 GPNVSD 162

RESULT 4
US-09-398-395A-54
; Sequence 54, Application US/09398395A
; Patent No. 6468772
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: No. 6468772, Joseph P.
; APPLICANT: Starks, Courtney M.
; APPLICANT: Manna, Kathleen R.
; TITLE OF INVENTION: SYNTHASES
; FILE REFERENCE: 07678-025001
; CURRENT APPLICATION NUMBER: US/09/398,395A
; CURRENT FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: 60/100,993
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/130,628
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: 60/150,262
; PRIOR FILING DATE: 1999-08-23
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
```

;; SEQ ID NO 54  
;; LENGTH: 590  
;; TYPE: PRT  
;; ORGANISM: Salvia officinalis  
US-09-398-39A-54

Query Match 1.4%; Score 7; DB 4; Length 590;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 397 GENVED 403  
Db 156 GENVED 162

RESULT 5  
US-08-937-931-8  
; Sequence 8, Application US/08937931  
; Patent No. 5935792  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerald M.  
; APPLICANT: Pan, Duojia  
; APPLICANT: Rooke, Jenny  
; APPLICANT: Yavari, Reza  
; APPLICANT: Xu, Tian  
; TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/937,931  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B97-081  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 749 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-937-931-8

Query Match 1.4%; Score 7; DB 2; Length 749;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AMPEKHA 117  
Db 196 AMPEKHA 202

RESULT 6  
US-09-285-502-8  
; Sequence 8, Application US/09285502  
; Patent No. 6190876  
; GENERAL INFORMATION:

;; APPLICANT: Rubin, Gerald M.  
;; APPLICANT: Pan, Duojia  
;; APPLICANT: Rooke, Jenny  
;; APPLICANT: Yavari, Reza  
;; APPLICANT: Xu, Tian  
;; TITLE OF INVENTION: KUZ: A No. 6190876el Family of Metalloproteases  
;; NUMBER OF SEQUENCES: 10  
;; CORRESPONDENCE ADDRESSES:  
;; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
;; STREET: 268 BUSH STREET, SUITE 3200  
;; CITY: SAN FRANCISCO  
;; STATE: CALIFORNIA  
;; COUNTRY: USA  
;; ZIP: 94104  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible

US-09-285-502-8

Query Match 1.4%; Score 7; DB 4; Length 749;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AMPEKHA 117  
Db 196 AMPEKHA 202

RESULT 7  
US-09-709-126-8  
; Sequence 8, Application US/09709126  
; Patent No. 6319704  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerald M.  
; APPLICANT: Pan, Duojia  
; APPLICANT: Rooke, Jenny  
; APPLICANT: Yavari, Reza  
; APPLICANT: Xu, Tian  
; TITLE OF INVENTION: KUZ: A No. 6319704el Family of Metalloproteases  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/709,126  
APPLICATION NUMBER: US/09/709,126  
FILING DATE: 08-NO. 6319704-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/285,502  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: B97-081  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 749 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-709-126-8

Query Match 1.4%; Score 7; DB 4; Length 749;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 AHPEKHA 117  
Db 196 AHPEKHA 202

RESULT 8  
US-09-871-385A-8  
Sequence 8, Application US/09871385A  
Patent No. 6399350  
GENERAL INFORMATION:  
APPLICANT: Rubin, Gerald M.  
Pan, DuoJia  
Rooke, Jenny  
Yavaril, Reza  
Xu, Tian  
TITLE OF INVENTION: KIT: A No. 6399350el Family of Metalloproteases  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 268 BUSH STREET, SUITE 3200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/871,385A  
FILING DATE: 31-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/709,126  
FILING DATE: 08-NO. 6399350-2000  
APPLICATION NUMBER: 09/285,502  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: B97-081  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 749 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-871-385A-8

Query Match 1.4%; Score 7; DB 4; Length 749;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 AHPEKHA 117  
Db 196 AHPEKHA 202

RESULT 9  
US-08-486-099-103  
Sequence 103 Application US/08486099  
Patent No. 6013263  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Danl P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Pelletway, Stephen R.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS  
B VIRUS TRANSMISSION  
TITLE OF INVENTION: B VIRUS TRANSMISSION  
NUMBER OF SEQUENCES: 209  
CORRESPONDENCE ADDRESS:  
ADDRESS: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/486,099  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-031  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 856 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
MOLECULE TYPE: protein  
US-08-486-099-103

Query Match 1.4%; Score 7; DB 3; Length 856;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194  
DB 842 PETAAL 848

## RESULT 10

US-08-484-223B-103  
Sequence 103, Application US/08484223B  
Patent No. 6020459  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
NUMBER OF SEQUENCES: 245  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,223B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Cortuzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 856 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-484-223B-103

Query Match 1.4%; Score 7; DB 3; Length 856;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194  
DB 842 PETAAL 848

## RESULT 11

US-08-919-597-103  
Sequence 103, Application US/08919597  
Patent No. 6054265  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.

APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/919,597  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/470,896  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Cortuzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 103:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 856 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-919-597-103

Query Match 1.4%; Score 7; DB 3; Length 856;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194  
DB 842 PETAAL 848

## RESULT 12

US-08-475-668A-103  
Sequence 103, Application US/08475668A  
Patent No. 6060065  
GENERAL INFORMATION:  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
NUMBER OF SEQUENCES: 211  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,668A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-026
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-475-668A-103

Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 856;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
Db 842 PETAAL 848

RESULT 13
US-08-485-551A-103
Sequence 103, Application US/08485551A
Patent No. 6068973
GENERAL INFORMATION:
APPLICANT: Bolognesi, Danl P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,551A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-023
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
```

```
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-551A-103

Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 856;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
Db 842 PETAAL 848

RESULT 14
US-08-471-913A-103
Sequence 103, Application US/08471913A
Patent No. 6093794
GENERAL INFORMATION:
APPLICANT: Bolognesi, Danl P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 214
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,913A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-030
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-471-913A-103

Query Match
Best Local Similarity 1.4%; Score 7; DB 3; Length 856;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 188 PETAAL 194
Db 842 PETAAL 848
```

DB 842 PETAAL 848

```

RESULT 15
US-08-485-264A-103
; Sequence 103, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Peteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 969-9741/8864
; TELEFAX: (212) 969-9741/8864
; TELEX: 66141 DENNIE
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 856 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-264A-103

Query Match 1.48; Score 7; DB 4; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 188 PETAAL 194
|1111111
DB 842 PETAAL 848

```

Search completed: November 29, 2002, 12:47:33  
 Job time : 25 secs



PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 40510  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC011159.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
US-09-864-761-40510  
Query Match 1.6%; Score 8; DB 10; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 23 HDDHDDDD 30  
DB 9 HDDHDDDD 16  
RESULT 2  
Sequence 3, Application US/08910386A  
Patent No. US20020092041A1  
GENERAL INFORMATION:  
APPLICANT: Ronald, Pamela C.  
APPLICANT: Wang, Guo-Liang  
APPLICANT: Song, Wen-Yuang  
APPLICANT: Hulbert, Scot  
APPLICANT: Richter, Todd  
TITLE OF INVENTION: Procedures and Materials for Confering  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,386A  
FILING DATE: 13-AUG-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-058950US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1445 amino acids

TYPE: amino acid  
TOPDOCK: linear  
MOLECULE TYPE: protein  
US-08-910-386A-3  
Query Match 1.6%; Score 8; DB 8; Length 1445;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 273 CDNLGARY 280  
DB 1372 CDNLGARY 1379  
RESULT 3  
US-09-764-868-863  
Sequence 863, Application US/09764868  
Patent No. US20020168711A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT232  
CURRENT APPLICATION NUMBER: US/09/764,868  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1510  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 863  
LENGTH: 192  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-868-863  
Query Match 1.4%; Score 7; DB 9; Length 192;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 470 LMARSLV 476  
DB 174 LMARSLV 180  
RESULT 4  
US-09-764-868-1175  
Sequence 1175, Application US/09764868  
Patent No. US20020168711A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT232  
CURRENT APPLICATION NUMBER: US/09/764,868  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1510  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1175  
LENGTH: 222  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (190)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (194)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (195)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (205)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-1175

```
Query Match 1.48; Score 7; DB 9; Length 222;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LNARSLV 476
|||||
DB 174 LNARSLV 180

RESULT 5
US-09-815-242-13488
; Sequence 13488, Application US/098152242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 14110
; SEQ ID NO 13488
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13488

Query Match 1.48; Score 7; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LNARSLV 476
|||||
DB 50 LNARSLV 56

RESULT 6
US-09-815-242-13556
; Sequence 13556, Application US/098152242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 14110
; SEQ ID NO 13556
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13556

Query Match 1.48; Score 7; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LNARSLV 476
|||||
DB 50 LNARSLV 56

RESULT 7
US-09-802-472B-2
; Sequence 2, Application US/09802472B
; Patent No. US2002010353A1
; GENERAL INFORMATION:
; APPLICANT: EINAT, Paz
; APPLICANT: SKALITER, Ramli
; APPLICANT: FEINSTEIN, Elena
; TITLE OF INVENTION: SEQUENCES CHARACTERISTIC OF HYPOXIA-REGULATED GENE TRANSCRIPTS
; FILE REFERENCE: EINAT-4.1C
; CURRENT APPLICATION NUMBER: US/09/802,472B
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: US 09/383,096
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: US 09/138,109
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: US 60/098,158
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: US 60/132,684
; PRIOR FILING DATE: 1999-05-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-802-472B-2

Query Match 1.48; Score 7; DB 10; Length 437;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 HGVNES 307
|||||
DB 354 HGVNES 360

RESULT 8
```

US-09-815-242-5114  
; Sequence 5114, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: Elinra, 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5114  
; LENGTH: 531  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5114  
Query Match 1.4%; Score 7; DB 10; Length 511;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 119 PGDLGDL 125  
Db 278 PGDLGDL 284

RESULT 9  
US-09-887-586A-54  
; Sequence 54, Application US/0987586A  
; Patent No. US20020094556A1  
; GENERAL INFORMATION:  
; APPLICANT: Chappell, Joseph  
; APPLICANT: No. US20020094556A1, Joseph P.  
; APPLICANT: Starks, Courtney M.  
; APPLICANT: Manna, Kathleen R.  
; TITLE OF INVENTION: SYNTHASES  
; FILE REFERENCE: 07678-025001  
; CURRENT APPLICATION NUMBER: US/09/887,586A  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: 09/398,395  
; PRIOR FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: 60/130,628  
; PRIOR FILING DATE: 1999-04-22  
; PRIOR APPLICATION NUMBER: 60/150,262  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 54  
; LENGTH: 590  
; TYPE: PRT  
; ORGANISM: Salvia officinalis

US-09-887-586A-54  
Query Match 1.4%; Score 7; DB 10; Length 590;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 397 GFNVSED 403  
Db 156 GFNVSED 162

RESULT 10  
US-09-903-012-54  
; Sequence 54, Application US/09903012  
; Patent No. US20020094557A1  
; GENERAL INFORMATION:  
; APPLICANT: Chappell, Joseph  
; APPLICANT: No. US20020094557A1, Joseph P.  
; APPLICANT: Starks, Courtney M.  
; APPLICANT: Manna, Kathleen R.  
; TITLE OF INVENTION: SYNTHASES  
; FILE REFERENCE: 07678-025001  
; CURRENT APPLICATION NUMBER: US/09/903,012  
; PRIOR FILING DATE: 2001-07-11  
; PRIOR APPLICATION NUMBER: 09/398,395  
; PRIOR FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: 60/100,993  
; PRIOR FILING DATE: 1998-09-16  
; PRIOR APPLICATION NUMBER: 60/130,628  
; PRIOR FILING DATE: 1999-04-22  
; PRIOR APPLICATION NUMBER: 60/150,262  
; PRIOR FILING DATE: 1999-08-23  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 54  
; LENGTH: 590  
; TYPE: PRT  
; ORGANISM: Salvia officinalis  
US-09-903-012-54  
Query Match 1.4%; Score 7; DB 10; Length 590;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 397 GFNVSED 403  
Db 156 GFNVSED 162

RESULT 11  
US-10-052-586-560  
; Sequence 560, Application US/10052586  
; Patent No. US20020127584A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C1  
; CURRENT APPLICATION NUMBER: US/10/052,586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266



us-09-868-760-7.oli.rapp

|    |                                      |
|----|--------------------------------------|
| 1  | PRIOR FILING DATE: 1997-09-18        |
| 2  | PRIOR APPLICATION NUMBER: 60/0662250 |
| 3  | PRIOR FILING DATE: 1997-10-17        |
| 4  | PRIOR APPLICATION NUMBER: 60/063120  |
| 5  | PRIOR FILING DATE: 1997-10-24        |
| 6  | PRIOR APPLICATION NUMBER: 60/063121  |
| 7  | PRIOR FILING DATE: 1997-10-24        |
| 8  | PRIOR APPLICATION NUMBER: 60/063486  |
| 9  | PRIOR FILING DATE: 1997-10-21        |
| 10 | PRIOR APPLICATION NUMBER: 60/063540  |
| 11 | PRIOR FILING DATE: 1997-10-28        |
| 12 | PRIOR APPLICATION NUMBER: 60/063541  |
| 13 | PRIOR FILING DATE: 1997-10-28        |
| 14 | PRIOR APPLICATION NUMBER: 60/063544  |
| 15 | PRIOR FILING DATE: 1997-10-28        |
| 16 | PRIOR APPLICATION NUMBER: 60/063564  |
| 17 | PRIOR FILING DATE: 1998-10-28        |
| 18 | PRIOR APPLICATION NUMBER: 60/063734  |
| 19 | PRIOR FILING DATE: 1998-10-29        |
| 20 | PRIOR APPLICATION NUMBER: 60/063870  |
| 21 | PRIOR FILING DATE: 1997-10-31        |
| 22 | PRIOR APPLICATION NUMBER: 60/064103  |
| 23 | PRIOR FILING DATE: 1997-10-30/065311 |
| 24 | PRIOR APPLICATION NUMBER: 60/0656120 |
| 25 | PRIOR FILING DATE: 1997-11-24        |
| 26 | PRIOR APPLICATION NUMBER: 60/066466  |
| 27 | PRIOR FILING DATE: 1997-11-24        |
| 28 | PRIOR APPLICATION NUMBER: 60/066772  |
| 29 | PRIOR FILING DATE: 1997-11-24        |
| 30 | PRIOR APPLICATION NUMBER: 60/069335  |
| 31 | PRIOR FILING DATE: 1997-12-11        |
| 32 | PRIOR APPLICATION NUMBER: 60/069425  |
| 33 | PRIOR FILING DATE: 1997-12-12        |
| 34 | PRIOR APPLICATION NUMBER: 60/069670  |
| 35 | PRIOR FILING DATE: 1997-12-17        |
| 36 | PRIOR APPLICATION NUMBER: 60/068017  |
| 37 | PRIOR FILING DATE: 1997-12-18        |
| 38 | PRIOR APPLICATION NUMBER: 60/077450  |
| 39 | PRIOR FILING DATE: 1998-03-10        |
| 40 | PRIOR APPLICATION NUMBER: 60/077632  |
| 41 | PRIOR FILING DATE: 1998-03-11        |
| 42 | PRIOR APPLICATION NUMBER: 60/077649  |
| 43 | PRIOR FILING DATE: 1998-03-11        |
| 44 | PRIOR APPLICATION NUMBER: 60/078886  |
| 45 | PRIOR FILING DATE: 1998-03-20        |
| 46 | PRIOR APPLICATION NUMBER: 60/078939  |
| 47 | PRIOR FILING DATE: 1998-03-20        |
| 48 | PRIOR APPLICATION NUMBER: 60/079664  |
| 49 | PRIOR FILING DATE: 1998-03-27        |
| 50 | PRIOR APPLICATION NUMBER: 60/079786  |
| 51 | PRIOR FILING DATE: 1998-03-27        |
| 52 | PRIOR APPLICATION NUMBER: 60/080107  |
| 53 | PRIOR FILING DATE: 1998-03-31        |
| 54 | PRIOR APPLICATION NUMBER: 60/080194  |
| 55 | PRIOR FILING DATE: 1998-03-31        |
| 56 | PRIOR APPLICATION NUMBER: 60/080327  |
| 57 | PRIOR FILING DATE: 1998-04-01        |
| 58 | PRIOR APPLICATION NUMBER: 60/080333  |
| 59 | PRIOR FILING DATE: 1998-04-01        |
| 60 | PRIOR APPLICATION NUMBER: 60/081049  |
| 61 | PRIOR FILING DATE: 1998-04-08        |
| 62 | PRIOR APPLICATION NUMBER: 60/081070  |
| 63 | PRIOR FILING DATE: 1998-04-08        |
| 64 | PRIOR APPLICATION NUMBER: 60/081195  |
| 65 | PRIOR FILING DATE: 1998-04-09        |
| 66 | PRIOR APPLICATION NUMBER: 60/081838  |
| 67 | PRIOR FILING DATE: 1998-04-15        |
| 68 | PRIOR APPLICATION NUMBER: 60/082568  |
| 69 | PRIOR FILING DATE: 1998-04-21        |
| 70 | PRIOR APPLICATION NUMBER: 60/082569  |
| 71 | PRIOR FILING DATE: 1998-04-21        |
| 72 | PRIOR APPLICATION NUMBER: 60/082569  |

1. PRIOR APPLICATION NUMBER: 60/082704  
2. PRIOR FILING DATE: 1998-04-22  
3. PRIOR APPLICATION NUMBER: 60/082757  
4. PRIOR FILING DATE: 1998-04-22  
5. PRIOR APPLICATION NUMBER: 60/083352  
6. PRIOR FILING DATE: 1998-04-28  
7. PRIOR APPLICATION NUMBER: 60/083495  
8. PRIOR FILING DATE: 1998-04-29  
9. PRIOR APPLICATION NUMBER: 60/083496  
10. PRIOR FILING DATE: 1998-04-29  
11. PRIOR APPLICATION NUMBER: 60/083499  
12. PRIOR FILING DATE: 1998-04-29  
13. PRIOR APPLICATION NUMBER: 60/083559  
14. PRIOR FILING DATE: 1998-04-29  
15. PRIOR APPLICATION NUMBER: 60/084366  
16. PRIOR FILING DATE: 1998-05-05  
17. PRIOR APPLICATION NUMBER: 60/084414  
18. PRIOR FILING DATE: 1998-05-06  
19. PRIOR APPLICATION NUMBER: 60/084639  
20. PRIOR FILING DATE: 1998-05-07  
21. PRIOR APPLICATION NUMBER: 60/084640  
22. PRIOR FILING DATE: 1998-05-07  
23. PRIOR APPLICATION NUMBER: 60/084643  
24. PRIOR FILING DATE: 1998-05-07  
25. PRIOR APPLICATION NUMBER: 60/085573  
26. PRIOR FILING DATE: 1998-05-15  
27. PRIOR APPLICATION NUMBER: 60/085579  
28. PRIOR FILING DATE: 1998-05-15  
29. PRIOR APPLICATION NUMBER: 60/085580  
30. PRIOR FILING DATE: 1998-05-15  
31. PRIOR APPLICATION NUMBER: 60/085582  
32. PRIOR FILING DATE: 1998-05-15  
33. PRIOR APPLICATION NUMBER: 60/085700  
34. PRIOR FILING DATE: 1998-05-15  
35. PRIOR APPLICATION NUMBER: 60/086023  
36. PRIOR FILING DATE: 1998-05-18  
37. PRIOR APPLICATION NUMBER: 60/086392  
38. PRIOR FILING DATE: 1998-05-22  
39. PRIOR APPLICATION NUMBER: 60/086486  
40. PRIOR FILING DATE: 1998-05-22  
41. PRIOR APPLICATION NUMBER: 60/087098  
42. PRIOR FILING DATE: 1998-05-28  
43. PRIOR APPLICATION NUMBER: 60/087208  
44. PRIOR FILING DATE: 1998-05-28  
45. PRIOR APPLICATION NUMBER: 60/087609  
46. PRIOR FILING DATE: 1998-06-02  
47. PRIOR APPLICATION NUMBER: 60/087759  
48. PRIOR FILING DATE: 1998-06-02  
49. PRIOR APPLICATION NUMBER: 60/087827  
50. PRIOR FILING DATE: 1998-06-03  
51. PRIOR APPLICATION NUMBER: 60/088025  
52. PRIOR FILING DATE: 1998-06-04  
53. PRIOR APPLICATION NUMBER: 60/088028  
54. PRIOR FILING DATE: 1998-06-04  
55. PRIOR APPLICATION NUMBER: 60/088029  
56. PRIOR FILING DATE: 1998-06-04  
57. PRIOR APPLICATION NUMBER: 60/088033  
58. PRIOR FILING DATE: 1998-06-04  
59. PRIOR APPLICATION NUMBER: 60/088167  
60. PRIOR FILING DATE: 1998-06-05  
61. PRIOR APPLICATION NUMBER: 60/088202  
62. PRIOR FILING DATE: 1998-06-05  
63. PRIOR APPLICATION NUMBER: 60/088212  
64. PRIOR FILING DATE: 1998-06-05  
65. PRIOR APPLICATION NUMBER: 60/088327  
66. PRIOR FILING DATE: 1998-06-05  
67. PRIOR APPLICATION NUMBER: 60/088326  
68. PRIOR FILING DATE: 1998-06-05  
69. PRIOR APPLICATION NUMBER: 60/088655  
70. PRIOR FILING DATE: 1998-06-09  
71. PRIOR APPLICATION NUMBER: 60/088722  
72. PRIOR FILING DATE: 1998-06-10  
73. PRIOR APPLICATION NUMBER: 60/088773

PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088740  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088811  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088824  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088825  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088826  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088861  
 PRIOR FILING DATE: 1998-06-11  
 PRIOR APPLICATION NUMBER: 60/088863  
 PRIOR FILING DATE: 1998-06-11  
 PRIOR APPLICATION NUMBER: 60/088876  
 PRIOR FILING DATE: 1998-06-11  
 PRIOR APPLICATION NUMBER: 60/089090  
 PRIOR FILING DATE: 1998-06-12  
 PRIOR APPLICATION NUMBER: 60/089105  
 PRIOR FILING DATE: 1998-06-12  
 PRIOR APPLICATION NUMBER: 60/089512  
 PRIOR FILING DATE: 1998-06-16  
 PRIOR APPLICATION NUMBER: 60/089514  
 PRIOR FILING DATE: 1998-06-16  
 PRIOR APPLICATION NUMBER: 60/089538  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/089598  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/089653  
 PRIOR FILING DATE: 1998-06-17  
 PRIOR APPLICATION NUMBER: 60/089908

Query Match 1.4% Score 7: DB 12: Length 692:  
 Best Local Similarity 100.0% Pred. No. 65:  
 Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 189 ETAAALH 195  
 DB 123 ETAAALH 129

RESULT 12  
 US-09-871-388-8  
 Sequence 8, Application US/09871388  
 Patent No. US20020127621A1  
 GENERAL INFORMATION:  
 APPLICANT: Rubin, Gerald M.  
 Pan, Duojia  
 Rookey, Jenny  
 Yavari, Reza  
 Xu, Tian  
 TITLE OF INVENTION: KUZ: A NO. US20020127621A1el Family of Metalloproteases  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: SCIENCE & TECHNOLOGY LAM GROUP  
 STREET: 268 BUSH STREET, SUITE 3200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/871,388  
 FILING DATE: 31-May-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/937,931  
 FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
 NAME: OSMAN, RICHARD A  
 REGISTRATION NUMBER: 36,627  
 REFERENCE/DOCKET NUMBER: B97-081  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 343-4341  
 TELEFAX: (415) 343-4342  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 749 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
 US-09-871-388-8

Query Match 1.4% Score 7: DB 10: Length 749:  
 Best Local Similarity 100.0% Pred. No. 70:  
 Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 AHPKHA 117  
 DB 196 AHPKHA 202

RESULT 13  
 US-09-815-242-13636  
 Sequence 13636, Application US/09815242  
 Patent No. US2002061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPINANT: Oulsen, Karl L.  
 APPINANT: Zysking, Judith W.  
 APPINANT: Wall, Dan John D.  
 APPINANT: Trawick, Grant J.  
 APPINANT: Car, Leo Robert T.  
 APPINANT: Yamano, Robert T.  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 PRIOR FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 13636  
 LENGTH: 810  
 TYPE: PRT  
 ORGANISM: Streptococcus pneumoniae  
 US-09-815-242-13636

Query Match 1.4% Score 7: DB 10: Length 810:  
 Best Local Similarity 100.0% Pred. No. 75:  
 Matches 7: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 158 SMTILOG 164  
 DB 374 SMTILOG 380

```
RESULT 14
US-09-881-752A-308
; Sequence 308, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 308
; LENGTH: 1797
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 85
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-881-752A-308

Query Match
Best Local Similarity 1.4%; Score 7; DB 10; Length 1797;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 INIFGDD 319
|||||
Db 276 INIFGDD 282

RESULT 15
US-09-815-242-11410
; Sequence 11410, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11410
; LENGTH: 2890
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11410

Query Match
Best Local Similarity 1.4%; Score 7; DB 10; Length 2890;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 313 INIFGDD 319
|||||
Db 1369 INIFGDD 1375

Search completed: November 29, 2002, 12:48:06
Job time : 19 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:42:30 ; Search time 20 Seconds

(without alignments)  
2388.940 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497

Sequence: 1 DGGCNDGONKDDHDDHDD.....MOGHEVESERVACVIGRA 497

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR-73.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID    | Description        |
|------------|-------|-------------|--------|----------|--------------------|
| 1          | 9     | 1.8         | 79     | 2 C84308 | hypothetical prote |
| 2          | 8     | 1.6         | 131    | 2 T25070 | hypothetical prote |
| 3          | 8     | 1.6         | 149    | 2 A54530 | eggshell protein - |
| 4          | 8     | 1.6         | 362    | 2 T08816 | choline kinase (EC |
| 5          | 8     | 1.6         | 378    | 2 AC0722 | probable cytochrom |
| 6          | 8     | 1.6         | 421    | 1 RK0ALG | ribulose-bisphosph |
| 7          | 8     | 1.6         | 421    | 2 S21984 | ribulose-bisphosph |
| 8          | 8     | 1.6         | 426    | 1 RKBHLC | ribulose-bisphosph |
| 9          | 8     | 1.6         | 441    | 1 RKBHLC | ribulose-bisphosph |
| 10         | 8     | 1.6         | 452    | 2 S47228 | ribulose-bisphosph |
| 11         | 8     | 1.6         | 452    | 2 S47229 | ribulose-bisphosph |
| 12         | 8     | 1.6         | 452    | 2 S47230 | ribulose-bisphosph |
| 13         | 8     | 1.6         | 452    | 2 S47231 | ribulose-bisphosph |
| 14         | 8     | 1.6         | 452    | 2 S47232 | ribulose-bisphosph |
| 15         | 8     | 1.6         | 452    | 2 S47233 | ribulose-bisphosph |
| 16         | 8     | 1.6         | 452    | 2 S47234 | ribulose-bisphosph |
| 17         | 8     | 1.6         | 452    | 2 S47235 | ribulose-bisphosph |
| 18         | 8     | 1.6         | 452    | 2 S47236 | ribulose-bisphosph |
| 19         | 8     | 1.6         | 452    | 2 S47237 | ribulose-bisphosph |
| 20         | 8     | 1.6         | 452    | 2 S47238 | ribulose-bisphosph |
| 21         | 8     | 1.6         | 452    | 2 S47239 | ribulose-bisphosph |
| 22         | 8     | 1.6         | 452    | 2 S47240 | ribulose-bisphosph |
| 23         | 8     | 1.6         | 452    | 2 S47241 | ribulose-bisphosph |
| 24         | 8     | 1.6         | 452    | 2 S47242 | ribulose-bisphosph |
| 25         | 8     | 1.6         | 452    | 2 S47243 | ribulose-bisphosph |
| 26         | 8     | 1.6         | 452    | 2 S47244 | ribulose-bisphosph |
| 27         | 8     | 1.6         | 452    | 2 S47245 | ribulose-bisphosph |
| 28         | 8     | 1.6         | 452    | 2 S47246 | ribulose-bisphosph |
| 29         | 8     | 1.6         | 458    | 2 T01639 | ribulose-bisphosph |

|    |   |     |     |          |                    |
|----|---|-----|-----|----------|--------------------|
| 30 | 8 | 1.6 | 469 | 1 RKBUL  | ribulose-bisphosph |
| 31 | 8 | 1.6 | 469 | 2 T01638 | ribulose-bisphosph |
| 32 | 8 | 1.6 | 469 | 2 T01640 | ribulose-bisphosph |
| 33 | 8 | 1.6 | 469 | 2 T01641 | ribulose-bisphosph |
| 34 | 8 | 1.6 | 469 | 2 T01644 | ribulose-bisphosph |
| 35 | 8 | 1.6 | 472 | 1 RKRRL2 | ribulose-bisphosph |
| 36 | 8 | 1.6 | 472 | 1 RKRCL  | ribulose-bisphosph |
| 37 | 8 | 1.6 | 472 | 1 RKRRL1 | ribulose-bisphosph |
| 38 | 8 | 1.6 | 472 | 2 T01642 | ribulose-bisphosph |
| 39 | 8 | 1.6 | 473 | 2 S34316 | ribulose-bisphosph |
| 40 | 8 | 1.6 | 473 | 2 S18315 | ribulose-bisphosph |
| 41 | 8 | 1.6 | 474 | 1 RKAALC | ribulose-bisphosph |
| 42 | 8 | 1.6 | 475 | 1 RKRHL  | ribulose-bisphosph |
| 43 | 8 | 1.6 | 475 | 1 RKRHL  | ribulose-bisphosph |
| 44 | 8 | 1.6 | 475 | 1 RKRHL  | ribulose-bisphosph |
| 45 | 8 | 1.6 | 475 | 1 RKRSLN | ribulose-bisphosph |

## ALIGNMENTS

RESULT 1  
C84308 hypothetical protein Vng1546h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence-revision 02-Feb-2001 #text-change 02-Feb-2001  
C:Accession: C84308  
R:Ng, M.Y.; Kennedy, S.P.; Mahatras, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.; Leithauser, B.; Keller, K.; Gruen, R.; Danon, M.J.; Hough, D.W.; Maddocks, D.G.; Ja, J.; K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;  
A>Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: AB4160; MUID:20504483; PMID:11016950  
A:Accession: C84308  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-79 <STO>  
A:Cross-references: GB:AE004437; MID:910581035; PIDN:AG19831.1; GSPDB:GN00138  
A:Genetics:  
A:Gene: VNG1546H

Query Match 1.8%; Score 9; DB 2; Length 79;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 LVDDDRGVY 133  
DB 55 LVDDDRGVY 63

## RESULT 2

T25070 hypothetical protein T21C9.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 15-Oct-1999  
C:Accession: T25070  
R:McMurray, A.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: T25070  
A:Accession: T25070  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-131 <MIL>  
A:Cross-references: EMBL:Z73098; PIDN:CAA9733.1; GSPDB:GN00023; CESP:T21C9.5  
A:Experimental source: clone T21C9  
C:Genetics:  
A:Gene: CESP:T21C9.5  
A:Map position: 5  
A:Introns: 24/2; 68/3

Query Match 1.6%; Score 8; DB 2; Length 131;  
Best Local Similarity 100.0%; Pred. No. 2.4;

|    | Matches | 8 ;      | Conservative | 0 ; | Mismatches | 0 ; | Indels | 0 ; | Gaps | 0 ; |
|----|---------|----------|--------------|-----|------------|-----|--------|-----|------|-----|
| QY | 431     | HGHDDAHD | 438          |     |            |     |        |     |      |     |
|    |         |          |              |     |            |     |        |     |      |     |
| Db | 124     | HGHDDAHD | 131          |     |            |     |        |     |      |     |

```

RESULT 3
A54530
eggshell protein - fluke (Schistosoma mansoni) (fragment)
C:Species: Schistosoma mansoni
C:Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 15-Oct-1994
C:Accession: A54530
R:Johnson, K.S.; Taylor, D.W.; Cordingley, J.S.
Mol. Biochem. Parasitol. 22, 89-100, 1987
A:Title: Possible eggshell protein gene from Schistosoma mansoni.
A:Reference number: A54530; NID:87115624; PMID:3100949
A:Accession: A54530
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-149 <J0H>
A:Cross-References: GB:M15371, NID:g160976; PID:g160977
A:Keywords: egg shell; tandem repeat

```

```

Query Match 1.6%; Score 8; DB 2; Length 149;
Best Local Similarity 100.0%;
Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 13 DHHDDHND 20
      |||||
Db 118 DHHDDHND 125

```

RESULT 4  
T08816  
chooline kinase (EC 2.7.1.32) CK2 - soybean  
C:Species: Glycine max (soybean)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T08816  
R:Dewey, R.E.; Monke, D.E.; Goode, J.H.  
Plant Physiol. 110, 1197-1205, 1996  
A:Title: Characterization of soybean choline kinase cDNAs and their expression in yeast  
A:Reference number: 216478; MUID:97088716; PMID:8934624  
A:Accession: T08816  
A:Status: Preliminary; translated from GB/EML/DBD  
A:Molecule type: mRNA  
A:Residues: 1-362 <DEW>  
A:Cross-references: EML:U43839, NID:g1438880, PIDN:AMC49375.1; PID:g1438881  
A:Experimental source: strain Dare  
C:Genetics;  
A:Gene: CK2  
C:Function:  
A:Description: catalyzes phosphorylation of choline to O-phosphocholine in the presence  
of keywords: phosphotransferase

```

Query Match      1.6%; Score 8; DB 2; Length 362;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      444 DLDGVDID 451
          |||||
Db      28 DLDGVDID 35

RESULT 5
AC0722
Probable cytochrome oxidase chain II [imported] - Salmonella enterica subsp. enterica serovar typhimurium
C:Species: Salmonella enterica subsp. enterica serovar typhimurium
A:Note: this species has also been called Salmonella typhimurium
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC0722
R:Perkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.;
T., T.; Connelton, F.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar

```

S.: Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
 A.; Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AC0722  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-378 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD05477.1; PID:g16502982; GSPDB:GN00176  
 C:Gene: STY1921  
 C:Superfamily: cytochrome d ubiquinol oxidase  
  
 Query Match 1.6%; Score 8; DB 2; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 6, 5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 6
RKOALG
Ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Aegilops squarrosa (fra
C:Species: Aegilops squarrosa
C:Date: 30-Sep-1992 #sequence_rev1sum 30-Sep-1992 #text_change 23-Mar-2001
R:Ogihara Y., Terachi T., Sasahuma, T.
A:submitted to the EMBL Data Library, August 1991
A:description: Molecular analysis of the hot spot region related to length mutations
region.
:Reference number: S17319

```

A:Cross-references: EMBL:X62119; NID:911324; PIDN:CA44038.1; PID:911325  
C:Comment: In addition to Lys-146, another lysine, it is not certain which, may be the  
C:Function:  
C:Biocatalytic: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate  
to form phosphoenolpyruvate and ribulose biphosphate large chain  
E:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase  
F:10.279/Active site: Lys (ribulose-bisphosphate-binding) #status predicted  
F:146/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status pre  
F:148/Binding site: magnesium (Asp) #status predicted

|                       |         |                |       |             |
|-----------------------|---------|----------------|-------|-------------|
| Query Match           | 1.6%;   | Score 8;       | DB 1; | Length 421; |
| Best Local Similarity | 100.0%; | Pred. No. 7.2; |       |             |
| Matches               | 8;      | Conservative   | 0;    | Mismatches  |
|                       |         |                | 0;    | Indels      |
|                       |         |                | 0;    | Gaps        |
| QY                    | 315     | IFGDDSVL       | 322   |             |
|                       |         |                |       |             |
| Db                    | 338     | IFGDDSVL       | 345   |             |

```

RESULT 7
S21984
Ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Aegilops crassa chlorop
C:Species: chloroplast Aegilops crassa
C:Date: 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 15-Jun-2001
C:Accession: S21984
R:Ogihara, Y.; Terachi, T.; Sasahuma, T.
Submitted to the EMBL Data Library, August 1991
A:Description: Molecular analysis of the hot spot region related to length mutations
Region
A:Reference number: S17319
A:Accession: S21984
A:Molecule type: DNA
A:Residues: 1421 <GGT>
A:Cross-references: EMBL:X62118; NID:g11308; PIDD:CAA44032.1; PID:g11309
C:Genetics:
C:Gene: rbcL

```

```

A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase;
F:120/279/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:148/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predi
F:148/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 421;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 338 IFGDDSVL 345

RESULT 8
RIBULC
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain precursor - barley chloroplast
C:Species: chloroplast Hordeum vulgare (barley)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C:Accession: S08610
R:Zurawski, G.; Cieseq, M.T.; Brown, A.H.D.
Genetics 106:735-749, 1984
A:Title: The nature of nucleotide sequence divergence between barley and maize chloroplast
A:Reference number: S07243
A:Accession: S08610
A:Molecule type: DNA
A:Residues: 1-426 <MAN>
A:Cross-references: EMBL:X00630; NID:911585; PIDN:CA5265.1; PID:911587
C:Comment: In addition to Lys-201, another lysine, it is not certain which, may be the s
C:Genetics:
A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase;
F:3-426/Product: ribulose-bisphosphate carboxylase large chain (fragment) #status predi
F:115/334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predi
F:203/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 426;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 393 IFGDDSVL 400

RESULT 9
RIBULC
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - wheat chloroplast (fragme
C:Species: chloroplast Triticum aestivum (common wheat)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Mar-2001
C:Accession: B33395; S17323
R:Houtz, R.L.; Stults, J.T.; Mulligan, R.M.; Tolbert, N.E.
Proc. Natl. Acad. Sci. U.S.A. 86, 1855-1859, 1989
A:Title: Post-translational modifications in the large subunit of ribulose bisphosphate
A:Reference number: A33395; MUID:89184526; PMID:2928307
A:Accession: B33395
A:Molecule type: protein
A:Residues: 1-19 <HOU>
R:Ogihara, Y.; Terachi, T.; Sasahuma, T.
submitted to the EMBL Data Library, August 1991
A:Description: Molecular analysis of the hot spot region related to length mutations in
region.
A:Reference number: S17319
A:Accession: S17323
A:Molecule type: DNA

```

```

A:Residues: 20-441 <OGI>
A:Cross-references: EMBL:X62117; NID:912343; PIDN:CA44027.1; PID:912344
C:Comment: In addition to Lys-165, another lysine, it is not certain which, may be th
C:Genetics:
A:Gene: rdCL
A:Genome: chloroplast
C:Function:
A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosph
C:Superfamily: ribulose-bisphosphate carboxylase large chain
C:Keywords: acetylated amino end; Calvin cycle; carbon dioxide fixation; carbon-carbo
F:1/Modified site: acetylated amino end (Pro) #status experimental
F:12/Modified site: N6,N6,N6-trimethyllysine (Lys) #status absent
F:139/298/Active site: Lys (ribulose-bisphosphate-binding) #status predicted
F:165/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status pre
F:167/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 1; Length 441;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 357 IFGDDSVL 364

RESULT 10
S47228
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium corsicum chloro
C:Species: chloroplast Gallium corsicum
C>Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
C:Accession: S47228; S39278
R:Manen, J.F.; Natali, A.
submitted to the EMBL Data Library, July 1994
A:Description: Comparison of the evolution of the rdCL sequence and of the atpB-rdCL
A:Reference number: S47221
A:Accession: S47228
A:Molecule type: DNA
A:Residues: 1-452 <MAN>
A:Cross-references: EMBL:X81096
R:Manen, J.F.; Natali, A.; Ehrendorfer, F.
submitted to the EMBL Data Library, December 1993
A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA in
F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status pre
F:203/Binding site: magnesium (Asp) #status predicted
Query Match
Best Local Similarity 1.6%; Score 8; DB 2; Length 452;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 315 IFGDDSVL 322
Db 393 IFGDDSVL 400

RESULT 11
S47229
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium elongatum chloro
C:Species: chloroplast Gallium elongatum
C>Date: 06-Jan-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999
C:Accession: S47229; S39279

```

R:Manen, J.F.; Natall, A.  
 Submitted to the EMBL Data Library, July 1994  
 A:Description: Comparison of the evolution of the rbcl sequence and of the atpB-rbcl not  
 A:Reference number: S47221  
 A:Accession: S47229  
 A:Molecule type: DNA  
 A:Residues: 1-452 <MAN1>  
 A:Cross-references: EMBL:X81098  
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.  
 Submitted to the EMBL Data Library, December 1993  
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA inter  
 A:Reference number: S39270  
 A:Accession: S39279  
 A:Molecule type: DNA  
 A:Residues: 1-52, 'X', 54-57 <MAN2>  
 A:Cross-references: EMBL:X76461; NID:9434869; PIDN:CA53999.1; PID:9434870  
 C:Genetics:  
 A:Gene: rbcl  
 A:Genome: chloroplast  
 C:Function:  
 A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate  
 C:Superfamily: ribulose-bisphosphate carboxylase large chain  
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c  
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted  
 F:201/Binding site: carbon dioxide (lys) (covalent) (by Rubisco activase) #status predic  
 F:203/Binding site: magnesium (asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322  
 |||||  
 Db 393 IFGDDSVL 400

RESULT 12

S47230  
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium lucidum chloroplast  
 C:Species: chloroplast Gallium lucidum  
 C:Date: 06-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 22-Jun-1999  
 C:Accession: S47230; S39280  
 R:Manen, J.F.; Natall, A.  
 Submitted to the EMBL Data Library, July 1994  
 A:Description: Comparison of the evolution of the rbcl sequence and of the atpB-rbcl not  
 A:Reference number: S47221  
 A:Accession: S47230  
 A:Molecule type: DNA  
 A:Residues: 1-452 <MAN1>  
 A:Cross-references: EMBL:X81100  
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.  
 Submitted to the EMBL Data Library, December 1993  
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA inter  
 A:Reference number: S39280  
 A:Accession: S39280  
 A:Molecule type: DNA  
 A:Residues: 1-52, 'X', 54-57 <MAN2>  
 A:Cross-references: EMBL:X76468; NID:9434871; PIDN:CA54006.1; PID:9434872  
 C:Genetics:  
 A:Gene: rbcl  
 A:Genome: chloroplast  
 C:Function:  
 A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate  
 C:Superfamily: ribulose-bisphosphate carboxylase large chain  
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c  
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted  
 F:201/Binding site: carbon dioxide (lys) (covalent) (by Rubisco activase) #status predic  
 F:203/Binding site: magnesium (asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322  
 |||||  
 Db 393 IFGDDSVL 400

RESULT 13

S47231  
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium palustre chlor  
 C:Species: chloroplast Gallium palustre  
 C:Date: 06-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 22-Jun-1999  
 C:Accession: S47231; S39283  
 R:Manen, J.F.; Natall, A.  
 Submitted to the EMBL Data Library, July 1994  
 A:Description: Comparison of the evolution of the rbcl sequence and of the atpB-rbcl  
 A:Reference number: S47221  
 A:Accession: S47231  
 A:Molecule type: DNA  
 A:Residues: 1-452 <MAN1>  
 A:Cross-references: EMBL:X81102  
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.  
 Submitted to the EMBL Data Library, December 1993  
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA in  
 A:Reference number: S39270  
 A:Accession: S39283  
 A:Molecule type: DNA  
 A:Residues: 1-52, 'X', 54-57 <MAN2>  
 A:Cross-references: EMBL:X76464; NID:9434877; PIDN:CA54002.1; PID:9434878  
 C:Genetics:  
 A:Gene: rbcl  
 A:Genome: chloroplast  
 C:Function:  
 A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosph  
 C:Superfamily: ribulose-bisphosphate carboxylase large chain  
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase  
 F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted  
 F:201/Binding site: carbon dioxide (lys) (covalent) (by Rubisco activase) #status pre  
 F:203/Binding site: magnesium (asp) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322  
 |||||  
 Db 393 IFGDDSVL 400

RESULT 14

S47232  
 ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium palustre chlorop  
 C:Species: chloroplast Gallium palustre  
 C:Date: 06-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 22-Jun-1999  
 C:Accession: S47232; S39281  
 R:Manen, J.F.; Natall, A.  
 Submitted to the EMBL Data Library, July 1994  
 A:Description: Comparison of the evolution of the rbcl sequence and of the atpB-rbcl  
 A:Reference number: S47221  
 A:Accession: S47232  
 A:Molecule type: DNA  
 A:Residues: 1-452 <MAN1>  
 A:Cross-references: EMBL:X81101  
 R:Manen, J.F.; Natall, A.; Ehrendorfer, F.  
 Submitted to the EMBL Data Library, December 1993  
 A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA in  
 A:Reference number: S39270  
 A:Accession: S39281  
 A:Molecule type: DNA  
 A:Residues: 1-52, 'X', 54-57 <MAN2>  
 A:Cross-references: EMBL:X76466; NID:9434873; PIDN:CA54004.1; PID:9434874  
 C:Genetics:  
 A:Gene: rbcl  
 A:Genome: chloroplast  
 C:Function:



A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate  
C:Superfamily: ribulose-bisphosphate carboxylase large chain  
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c  
F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted  
F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predic  
F:203/Binding site: magnesium (Asp) #status predicted

Query Match 1.6% Score 8: DB 2: Length 452;  
Best Local Similarity 100.0%; Pred.No. 7.7;  
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322  
DB 393 IFGDDSVL 400

# RESULT 15

S47233  
ribulose-bisphosphate carboxylase (EC 4.1.1.39) large chain - Gallium scabrum chloroplast  
C:Species: chloroplast Gallium scabrum  
C:Date: 06-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 22-Jun-1999  
C:Accession: S47233; S39284  
R:Manen, J.F.; Natall, A.  
Submitted to the EMBL Data Library, July 1994  
A:Description: Comparison of the evolution of the rbcl sequence and of the atpB-rbcl non  
A:Reference number: S47221

A:Accession: S47233  
A:Molecule type: DNA  
A:Residues: 1452 <MAN1>  
A:Cross-references: EMBL:X81105  
R:Manen, J.F.; Natall, A.; Ehendorfer, F.  
Submitted to the EMBL Data Library, December 1993  
A:Description: Phylogeny of Rubiaceae-Rubia inferred from the sequence of a cpDNA inter  
A:Reference number: S39270

A:Accession: S39284  
A:Molecule type: DNA  
A:Residues: 1-52, 'X', 54-57 <MAN2>  
A:Cross-references: EMBL:X76462; NID:9434879; PIDN:CA54000.1; PID:9434880  
C:Genetics:  
A:Gene: rbcl  
A:Genome: chloroplast  
C:Function:

A:Description: catalyzes the reaction of carbon dioxide with D-ribulose 1,5-bisphosphate  
C:Superfamily: ribulose-bisphosphate carboxylase large chain  
C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carboxy-lyase; c  
F:175,334/Active site: Lys (ribulose-bisphosphate-binding) #status predicted  
F:201/Binding site: carbon dioxide (Lys) (covalent) (by Rubisco activase) #status predic  
F:203/Binding site: magnesium (Asp) #status predicted

Query Match 1.6% Score 8: DB 2: Length 452;  
Best Local Similarity 100.0%; Pred.No. 7.7;  
Matches 8: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDDSVL 322  
DB 393 IFGDDSVL 400

Search completed: November 29, 2002, 12:46:52  
Job time : 24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:37:40 ; Search time 17 Seconds

(without alignments)  
1212.573 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497  
Sequence: 1 DGEQNDGQNKDDHDDHD.....MOGHEVESERVACCVIGRA 497

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size: 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID           | Description         |
|------------|-------|-------------|--------|--------------|---------------------|
| 1          | 9     | 1.8         | 471    | 1 RBL2_HYDMR | Q59460 hydrogenvi   |
| 2          | 8     | 1.6         | 25     | 1 R706_BOVIN | P82931 bos taurus   |
| 3          | 8     | 1.6         | 125    | 1 R706_MOUSE | P58064 mus musculu  |
| 4          | 8     | 1.6         | 149    | 1 ECGS-SCHMA | P08016 schistosoma  |
| 5          | 8     | 1.6         | 394    | 1 RBL_ALIPY  | P34767 allisma plan |
| 6          | 8     | 1.6         | 394    | 1 RBL_BARLO  | 005795 barclaya lo  |
| 7          | 8     | 1.6         | 394    | 1 RBL_BRASC  | 005796 brasenia sc  |
| 8          | 8     | 1.6         | 394    | 1 RBL_CABCA  | 005797 cabomba car  |
| 9          | 8     | 1.6         | 394    | 1 RBL_CERDE  | 005798 ceratophyll  |
| 10         | 8     | 1.6         | 394    | 1 RBL_EURPE  | 005799 euryale fer  |
| 11         | 8     | 1.6         | 394    | 1 RBL_NELIU  | 005800 nelumbo lut  |
| 12         | 8     | 1.6         | 394    | 1 RBL_NYMPA  | 005801 nymphea vari |
| 13         | 8     | 1.6         | 394    | 1 RBL_NYMPD  | 005802 nymphea od   |
| 14         | 8     | 1.6         | 394    | 1 RBL_VICZ   | 005803 victoria cr  |
| 15         | 8     | 1.6         | 406    | 1 RBL_CREBT  | P48705 cleptodane   |
| 16         | 8     | 1.6         | 410    | 1 RBL_GLEJA  | P43223 gleichenia   |
| 17         | 8     | 1.6         | 413    | 1 RBL_NDIPE  | P43223 adiantum pe  |
| 18         | 8     | 1.6         | 414    | 1 RBL_RUEOR  | P43226 blechnum pe  |
| 19         | 8     | 1.6         | 414    | 1 RBL_ONIYA  | Q36510 onychium ja  |
| 20         | 8     | 1.6         | 415    | 1 RBL_CIBBA  | P43228 onychium ba  |
| 21         | 8     | 1.6         | 416    | 1 RBL_ARTBA  | P43228 arthropodi   |
| 22         | 8     | 1.6         | 416    | 1 RBL_SPIRK  | P36488 spigelia ma  |
| 23         | 8     | 1.6         | 420    | 1 RBL_ANEME  | Q31674 anemina mexi |
| 24         | 8     | 1.6         | 421    | 1 RBL_AEGCR  | P25114 aegilops cr  |
| 25         | 8     | 1.6         | 421    | 1 RBL_AEGTA  | P25114 aegilops ta  |
| 26         | 8     | 1.6         | 422    | 1 OMD_BOVIN  | Q77742 bos taurus   |
| 27         | 8     | 1.6         | 426    | 1 RBL_HORVU  | P05698 hordeum vul  |
| 28         | 8     | 1.6         | 435    | 1 RBL_DROBR  | P28404 drosera bur  |
| 29         | 8     | 1.6         | 436    | 1 RBL_SANCA  | P28450 sanguiaria   |
| 30         | 8     | 1.6         | 440    | 1 RBL_BAMGL  | P51994 bambusa gla  |
| 31         | 8     | 1.6         | 440    | 1 RBL_DICAN  | P48701 dicksonia a  |
| 32         | 8     | 1.6         | 440    | 1 RBL_MATST  | P48707 matuccia a   |
| 33         | 8     | 1.6         | 440    | 1 RBL_ONOSE  | P48710 onoclea sen  |

|    |   |     |     |             |                      |
|----|---|-----|-----|-------------|----------------------|
| 34 | 8 | 1.6 | 440 | 1 RBL_POLMU | P48712 polystichum   |
| 35 | 8 | 1.6 | 440 | 1 RBL_PTEVI | Q33015 pteris vitt   |
| 36 | 8 | 1.6 | 441 | 1 RBL_APOCA | Q05984 apocynum ca   |
| 37 | 8 | 1.6 | 441 | 1 RBL_ASCCX | P36480 asclepias e   |
| 38 | 8 | 1.6 | 441 | 1 RBL_BECMS | P28383 begonia met   |
| 39 | 8 | 1.6 | 441 | 1 RBL_CHELA | Q32016 cheilanthes   |
| 40 | 8 | 1.6 | 441 | 1 RBL_CORSA | Q05579 coriandrum    |
| 41 | 8 | 1.6 | 441 | 1 RBL_DARCA | P28398 dactylis glom |
| 42 | 8 | 1.6 | 441 | 1 RBL_DROPC | P28406 drosera dic   |
| 43 | 8 | 1.6 | 441 | 1 RBL_DROPT | P28410 drosera pet   |
| 44 | 8 | 1.6 | 441 | 1 RBL_FOUPP | Q05990 fouquieria    |
| 45 | 8 | 1.6 | 441 | 1 RBL_GLYEC | Q62570 glycyrhiza    |

## ALIGNMENTS

```

RESULT 1
ID RBL2_HYDMR STANDARD: PRT: 471 AA.
AC Q59460;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ribulose biphosphate carboxylase large chain 2 (EC 4.1.1.39)
DE (Rubisco large subunit).
DE CBL-2.
OS Hydrogenovibrio marinus.
OC Bacteria; Proteobacteria; gamma subdivision; Piscirickettsia group;
OC Hydrogenovibrio.
OX NCBI_TaxID=28885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH-110;
RA Yaguchi T., Chung S., Yasuo I., Tohru K.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-Ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-Ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: DA3622; BAA07731.1;
CC DR HSSP: P09657; IERN.
CC DR InterPro: IPR000685; Rubisco_large.
CC DR Pfam: PF00016; Rubisco_large; 1.
CC DR Pfam: PF02788; Rubisco_large; N. 1.
CC DR PROSITE: PS00157; RUBISCO_LARGE; FALSE_NEG.
CC KW Carbon dioxide fixation; Photorespiration; Lyase; Oxidoreductase;
CC FT Monooxygenase; Calvin cycle; Multigene family.
CC FT ACT_SITE 194
CC FT BINDING OF CO(2) ACTIVATES THE ENZYME
CC FT (BY SIMILARITY).
CC SQ SEQUENCE 471 AA: 52005 MW; BAD462E46451AD50 CRC64;

Query Match 1.8%; Score 9; DB 1; Length 471;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 314 NIFGDDSVL 322  
 |||||  
 DB 385 NIFGDDSVL 393

## RESULT 2

RT06\_BOVIN STANDARD: PRT: 25 AA.  
 AC P82931;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mitochondrial 28S ribosomal protein S6 (MRP-S6) (Fragments).  
 GN MRPS6 OR RPSM6.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 ON NCBI\_TaxID=9913;  
 OX [1]  
 RP SEQUENCE:  
 RC TISSUE=Liver;  
 RA MEDLINE=21276436; PubMed=11279123;  
 RA Koc E.C.; Burkhardt W.; Blackburn K.; Moseley A.; Spremulli L.L.;  
 RT "The small subunit of the mammalian mitochondrial ribosome:  
 RT Identification of the full complement of ribosomal proteins present."  
 RL J. Biol. Chem. 276:19363-19374(2001).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -1- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.  
 DR InterPro: IPR000529; Ribosomal\_S6.  
 DR Pfam: PF01250; Ribosomal\_S6; PARTIAL.  
 DR PROSITE: PS01048; RIBOSOMAL\_S6; PARTIAL.  
 KW Ribosomal protein; Mitochondrion.  
 FT NON\_TER 1 1  
 FT NON\_CONS 12 13  
 FT NON\_TER 25 25  
 SQ SEQUENCE 25 AA: 2710 MW; 57C89F522C889970 CRC64;  
 Query Match 1.6%; Score 8; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 187 RPERAAL 194  
 |||||  
 DB 4 RPERAAL 11

## RESULT 3

RT06\_MOUSE STANDARD: PRT: 125 AA.  
 AC P58064;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Mitochondrial 28S ribosomal protein S6 (MRP-S6).  
 GN MRPS6 OR RPSM6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RA SUZUKI T., TERASAKI M., TAKEMOTO C., HANADA T., WADA A., UEDA T.,  
 RA Watanabe K.;  
 RT "Mammalian mitochondrial ribosome: structural and functional  
 RT compensation for deficit of RNA with protein components."  
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -1- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AB049943; BAB40996.1; ALT\_INIT.  
 DR InterPro: IPR000529; Ribosomal\_S6.  
 DR TIGRFAMs: TIGR00166; S6; 1.  
 DR PROSITE: PS01048; RIBOSOMAL\_S6; FALSE\_NEG.  
 KW Ribosomal protein; Mitochondrion.  
 SQ SEQUENCE 125 AA: 14308 MW; 8140DB0BFPDD9025 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 187 RPERAAL 194  
 |||||  
 DB 15 RPERAAL 22

## RESULT 4

EGGS\_SCHEMA STANDARD: PRT: 149 AA.  
 ID EGGS\_SCHEMA  
 AC P08016;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Putative eggshell protein (Fragment).  
 OS Schistosoma mansoni (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigoididae;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 ON NCBI\_TaxID=6183;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87115624; PubMed=3100949;  
 RA Johnson K.S.; Taylor D.W.; Cordingley J.S.;  
 RA "Possible eggshell protein gene from Schistosoma mansoni."  
 RL Mol. Biochem. Parasitol. 22:89-100(1987).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M15371; AAA29877.1; .  
 DR PIR: A54530; A54530.  
 KW Eggshell; Repeat.  
 FT NON\_TER 1 1  
 FT DOMAIN 109 145 HIS-RICH.  
 SQ SEQUENCE 149 AA: 18839 MW; F5399D570845274A CRC64;

Query Match 1.6%; Score 8; DB 1; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 DHHDHHD 20  
 |||||  
 DB 118 DHHDHHD 125

## RESULT 5

RBL\_ALIPL STANDARD: PRT: 394 AA.  
 ID RBL\_ALIPL  
 AC P34767;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO

```
DE large subunit) (Fragment).
GN RBCL.
OS Alisma plantago-aquatica (Damasonium).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Alismaceae; Alisma.
ON NCBI_TaxID=15000;
RN [1]
RP SEQUENCE FROM N.A.
RA Chase M.W., Solitis D.E., Olmstead R.G., Morgan D., Les D.H.,
RA Mishler B.D., Duvall M.R., Price R.A., Hills H.G., Qiu Y.L.,
RA Kron K.A., Reel J.H., Conti E., Palmer J.D., Manhart J.R.,
RA Sydes K.J., Michaels H.J., Kress W.J., Karol K.G., Clark W.D.,
RA Hedren M., Gaut B.S., Jansen R.K., Kim K.J., Wimpee C.F., Smith J.F.,
RA Furler G.R., Strauss S.H., Xiang Q.-Y., Plunkett G.M., Solitis P.S.,
RA Swensen S., Williams S.E., Gadek P.A., Quinn C.J., Eguliste L.E.,
RA Golenberg E., Leach G.H., Graham S.W., Barrett S.C.H., Dayanandan S.,
RA Albert V.A.;
RT "Phylogenetics of seed plants: an analysis of nucleotide sequences
RT from the plastid gene rbcL.";
RL Ann. Mo. Bot. Gard. 80:528-580(1993).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATORY PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-
CC phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08759; AAA85294.1;
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large_N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43967 MW; 07C2DA05D1A322C2 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best local similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IRGDSVYL 322
DB 384 IRGDSVYL 391

RESULT 6
RBL_BARLO
ID RBL_BARLO STANDARD: PRT; 394 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
```

```
GN RBCL.
OS Barcleya longifolia.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Nymphaeaceae; Barcleya.
ON NCBI_TaxID=4412;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-92052222; PubMed-1946432;
RA Les D.H., Garvin D.K., Wimpee C.F.;
RT "Molecular evolutionary history of ancient aquatic angiosperms.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATORY PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-
CC phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77028; AAA84035.1;
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large; 1.
DR Pfam: PF02788; Rubisco_large_N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43804 MW; EF9671830EF55132 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best local similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IRGDSVYL 322
DB 384 IRGDSVYL 391

RESULT 7
RBL_BRASC
ID RBL_BRASC STANDARD: PRT; 394 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBCL.
OS Brasenia schreberi (Water shield).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Cabombaceae; Brasenia.
ON NCBI_TaxID=4424;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-92052222; PubMed-1946432;
```

RA Les D.H., Garvin D.K., Wimpsee C.F.;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) =  
CC D-RIBULOSE: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC CARBON DIOXIDE FIXATION, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-  
CC phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =  
CC 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77031; AAA84085.1; -  
DR InterPro: IPR000685; Rubisco\_large.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR Pfam: PF02788; Rubisco\_large; N: 1.  
DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 1  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394  
SQ SEQUENCE 394 AA: 43778 MW: 36DB85FE4408082F CRC64;  
  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 315 IFGDSVYL 322  
DB 384 IFGDSVYL 391  
  
RESULT 8  
RBL\_CARBA STANDARD; PRT: 394 AA.  
ID RBL\_CARBA  
AC Q05797; 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO  
DE large subunit) (Fragment).  
GN RBCL.  
OS Cabomba caroliniana.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Cabombaceae; Cabomba.  
OX NCBI\_TaxID=4426;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Les D.H., Garvin D.K., Wimpsee C.F.;  
RX MEDLINE=92052222; PubMed=1946432;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -----

CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-  
CC phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =  
CC 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77027; AAA84098.1; -  
DR InterPro: IPR000685; Rubisco\_large.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR Pfam: PF02788; Rubisco\_large; N: 1.  
DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 1  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394  
SQ SEQUENCE 394 AA: 43810 MW: B747E1471771EBB3 CRC64;  
  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 315 IFGDSVYL 322  
DB 384 IFGDSVYL 391  
  
RESULT 9  
RBL\_CERDE STANDARD; PRT: 394 AA.  
ID RBL\_CERDE  
AC Q05798; 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO  
DE large subunit) (Fragment).  
GN RBCL.  
OS Ceratophyllum demersum.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Ceratophyllales; Ceratophyllaceae;  
OC Ceratophyllum.  
OX NCBI\_TaxID=4428;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Les D.H., Garvin D.K., Wimpsee C.F.;  
RX MEDLINE=92052222; PubMed=1946432;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-  
CC phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =  
CC 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77030: AAA84104.1: -  
DR InterPro: IPR000685: Rubisco\_large.  
DR Pfam: PF00016: Rubisco\_large; 1.  
DR Pfam: PF02788: Rubisco\_large; N: 1.  
DR PROSITE: PS00157: RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 192  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394  
SQ SEQUENCE 394 AA: 43948 MW: 700780498919405 CRC64;  
  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 315 IFGDSVL 322  
Db 384 IFGDSVL 391  
| | | | | | | | | |  
RESULT 10  
RBL\_EURPE STANDARD: PRT; 394 AA.  
AC 005799:  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco  
DE large subunit) (Fragment).  
GN RBCL.  
OS Euryale ferox.  
OC Chloroplast.  
CC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; Nymphaeaceae; Euryale.  
CX NCBI\_TaxID:4414;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:92052222; PubMed:1946432;  
RA Les D.H., Garvin D.K., Wimpsee C.F.;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-  
CC phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) ->  
CC 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77035: AAA84249.1: -  
DR InterPro: IPR000685: Rubisco\_large.  
DR Pfam: PF00016: Rubisco\_large; 1.  
DR Pfam: PF02788: Rubisco\_large; N: 1.  
DR PROSITE: PS00157: RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 192  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394

DR InterPro: IPR000685: Rubisco\_large.  
DR Pfam: PF00016: Rubisco\_large; 1.  
DR Pfam: PF02788: Rubisco\_large; N: 1.  
DR PROSITE: PS00157: RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 192  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394  
SQ SEQUENCE 394 AA: 43862 MW: 3D9FA4821BF4143D CRC64;  
  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 315 IFGDSVL 322  
Db 384 IFGDSVL 391  
| | | | | | | | | |  
RESULT 11  
RBL\_NELLU STANDARD: PRT; 394 AA.  
AC 005800:  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco  
DE large subunit) (Fragment).  
GN RBCL.  
OS Nelumbo lutea (American lotus).  
OC Chloroplast.  
CC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; eudicotyledons; Nelumbonaceae; Nelumbo.  
CX NCBI\_TaxID:4431;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE:92052222; PubMed:1946432;  
RA Les D.H., Garvin D.K., Wimpsee C.F.;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF  
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
CC ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-  
CC phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) ->  
CC 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77032: AAA84499.1: -  
DR InterPro: IPR000685: Rubisco\_large.  
DR Pfam: PF00016: Rubisco\_large; 1.  
DR Pfam: PF02788: Rubisco\_large; N: 1.  
DR PROSITE: PS00157: RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1 192  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394

SO SEQUENCE 394 AA: 43917 MW: 93295743D36D2E4E CRC64;  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 315 IFGDDSVL 322  
|111111|  
DB 384 IFGDDSVL 391  
RESULT 12  
RBL\_NUPVA STANDARD: PRT: 394 AA.  
ID RBL\_NUPVA  
AC 005801;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco large subunit) (Fragment).  
GN RBCL.  
OS Nuphar variegata (Yellow pond lily).  
OC Chloroplast.  
CC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; Nymphaeaceae; Nuphar.  
CX NCBI\_TaxID=4416;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=92052222; PubMed=1946432;  
RA Les D.H., Garvin D.K., Winsee C.F.;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -> 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77029; AAA84527.1; -;  
DR InterPro: IPR000685; Rubisco\_large.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR Pfam: PF02788; Rubisco\_large.N; 1.  
DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394  
SQ SEQUENCE 394 AA: 43816 MW: E400510ADB98C803 CRC64;  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 315 IFGDDSVL 322  
|111111|  
DB 384 IFGDDSVL 391

RESULT 13  
RBL\_NYMOD STANDARD: PRT: 394 AA.  
ID RBL\_NYMOD  
AC 005802;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco large subunit) (Fragment).  
GN RBCL.  
OS Nymphaea odorata (White water lily).  
OC Chloroplast.  
CC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; Nymphaeaceae; Nymphaea.  
CX NCBI\_TaxID=4419;  
[1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=92052222; PubMed=1946432;  
RA Les D.H., Garvin D.K., Winsee C.F.;  
RT "Molecular evolutionary history of ancient aquatic angiosperms.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10119-10123(1991).  
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) -> 2 3-phospho-D-glycerate.  
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) -> 3-phospho-D-glycerate + 2-phosphoglycolate.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M77034; AAA84529.1; -;  
DR InterPro: IPR000685; Rubisco\_large.  
DR Pfam: PF00016; Rubisco\_large; 1.  
DR Pfam: PF02788; Rubisco\_large.N; 1.  
DR PROSITE: PS00157; RUBISCO\_LARGE; 1.  
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;  
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.  
FT NON\_TER 1  
FT ACT\_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.  
FT NON\_TER 394 394  
SQ SEQUENCE 394 AA: 43832 MW: 31E3C25FBBD0A7EF CRC64;  
Query Match 1.6%; Score 8; DB 1; Length 394;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 315 IFGDDSVL 322  
|111111|  
DB 384 IFGDDSVL 391  
RESULT 14  
RBL\_VYCCZ STANDARD: PRT: 394 AA.  
ID RBL\_VYCCZ  
AC 005803;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)



```

DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBCL.
OS Victoria cruziana (Santa Cruz water lily).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Nymphaeaceae; Victoria.
ON NCBI_TaxID=4421;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92052222; PubMed=1946432;
RA Les D.H., Garin D.K., Wimpsee C.F.;
RT "Molecular evolutionary history of ancient aquatic angiosperms.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:10119-10123(1991).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77036; AA84716.1; -.
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
DR Pfam: PF02788; RUBISCO_large_N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 192 192 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 394 394
SQ SEQUENCE 394 AA; 43848 MW; 13EF14825FEAE13D CRC64;

Query Match 1.6%; Score 8; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 2; 9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IRGDDSVL 322
DB 384 IRGDDSVL 391

```

```

RP SEQUENCE FROM N.A.
RC Tissue-Lesif;
RX MEDLINE=94261660; PubMed=8202555;
RA Hasebe M., Omori T., Nakazawa M., Sano T., Kato M., Iwatsuki K.;
RT "rbcL gene sequences provide evidence for the evolutionary lineages
RT of leptosporangiate ferns.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5730-5734(1994).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-
CC phospho-D-glycerate.
CC -1- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) =
CC 3-phospho-D-glycerate + 2-phosphoglycolate.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U05613; AAA1897.1; -.
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large; 1.
DR Pfam: PF02788; RUBISCO_large_N; 1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 179 179 BINDING OF CO(2) ACTIVATES THE ENZYME.
FT NON_TER 406 406
SQ SEQUENCE 406 AA; 44888 MW; 47137F9AED3F7950 CRC64;

Query Match 1.6%; Score 8; DB 1; Length 406;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IRGDDSVL 322
DB 371 IRGDDSVL 378

```

Search completed: November 29, 2002, 12:45:18  
 Job time : 20 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:38:35 : Search time 32 Seconds  
(without alignment)  
3200.169 Million cell updates/sec

Title: US-09-868-760-7

Sequence: 1 DGECCNDGQNKDHHDDHDD.....MOGHEVESERVACVIGRA 497

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: SPTREMBL\_21:\*

1: sp-archaea:\*

2: sp-bacteria:\*

3: sp-fungi:\*

4: sp-human:\*

5: sp-invertebrate:\*

6: sp-mammal:\*

7: sp-mhc:\*

8: sp-organelle:\*

9: sp-phage:\*

10: sp-plant:\*

11: sp-proteic:\*

12: sp-virus:\*

13: sp-vertebrate:\*

14: sp-unclassified:\*

15: sp-virus:\*

16: sp-bacteriap:\*

17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 434   | 87.3        | 517    | 5     | Q9BKB9      |
| 2          | 9     | 1.8         | 20     | 5     | P83148      |
| 3          | 9     | 1.8         | 79     | 17    | Q9HPN6      |
| 4          | 9     | 1.8         | 473    | 2     | O85040      |
| 5          | 9     | 1.8         | 473    | 2     | O85040      |
| 6          | 8     | 1.6         | 83     | 2     | O86380      |
| 7          | 8     | 1.6         | 83     | 2     | O86380      |
| 8          | 8     | 1.6         | 83     | 2     | O86380      |
| 9          | 8     | 1.6         | 83     | 2     | O86380      |
| 10         | 8     | 1.6         | 83     | 2     | O86380      |
| 11         | 8     | 1.6         | 83     | 2     | O86380      |
| 12         | 8     | 1.6         | 83     | 2     | O86380      |
| 13         | 8     | 1.6         | 83     | 2     | O86380      |
| 14         | 8     | 1.6         | 83     | 2     | O86380      |
| 15         | 8     | 1.6         | 83     | 2     | O86380      |
| 16         | 8     | 1.6         | 83     | 2     | O86380      |

|    |   |     |    |   |        |                    |
|----|---|-----|----|---|--------|--------------------|
| 17 | 8 | 1.6 | 83 | 2 | Q939W0 | Q939W0 anabaena so |
| 18 | 8 | 1.6 | 83 | 2 | Q939V5 | Q939V5 anabaena sp |
| 19 | 8 | 1.6 | 83 | 2 | Q939V4 | Q939V4 anabaena sp |
| 20 | 8 | 1.6 | 83 | 2 | Q939V3 | Q939V3 anabaena sp |
| 21 | 8 | 1.6 | 83 | 2 | Q939V2 | Q939V2 anabaena sp |
| 22 | 8 | 1.6 | 83 | 2 | Q939V1 | Q939V1 anabaena sp |
| 23 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 24 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 25 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 26 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 27 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 28 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 29 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 30 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 31 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 32 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 33 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 34 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 35 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 36 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 37 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 38 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 39 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 40 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 41 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 42 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 43 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 44 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |
| 45 | 8 | 1.6 | 83 | 2 | Q939V0 | Q939V0 anabaena sp |

## ALIGNMENTS

## RESULT 1

| ID | Q9BKB9  | PRELIMINARY: | PRT:                   | 517 AA. |
|----|---|--------------|------------------------|---------|
| AC | Q9BKB9  |              |                        |         |
| DT | 01-JUN-2001 (TREMBL)  | 17           | Created                |         |
| DT | 01-JUN-2001 (TREMBL)  | 17           | Last sequence update   |         |
| DT | 01-JUN-2002 (TREMBL)  | 21           | Last annotation update |         |
| DE | Pernin precursor.   |              |                        |         |
| OS | Pernin canaliculus (greenshell mussel).                               |              |                        |         |
| OC | Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;     |              |                        |         |
| OC | Mytiloidea; Mytilidae; Perna.   |              |                        |         |
| OX | NCBI_Taxid=38949;   |              |                        |         |
| RN | [1]   |              |                        |         |
| RP | SEQUENCE FROM N.A.  |              |                        |         |
| RX | MEDLINE=21186417; PubMed=11290459;                                    |              |                        |         |
| RA | Scotti P.D., Dearing S.C., Greenwood D.R., Newcomb R.D.;              |              |                        |         |
| RT | "Pernin: a novel self-aggregating haemolymph protein from the New     |              |                        |         |
| RT | Zealand green-lipped mussel Perna canaliculus (bivalvia: mytilidae)." |              |                        |         |
| RL | Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 128:767-779(2001).     |              |                        |         |
| CC | - FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  |              |                        |         |
| CC | CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).            |              |                        |         |
| CC | - CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).        |              |                        |         |
| CC | - SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.       |              |                        |         |
| DR | EMBL: AF273766; AAK20952.1; -   |              |                        |         |
| DR | HSSP: P00445; ITCV.   |              |                        |         |
| DR | InterPro: IPR001424; SOD_CU_ZN.                                       |              |                        |         |
| DR | Pfam: PF00080; sodcu.3  |              |                        |         |
| DR | PRINTS: PR00068; CUZNDISMUTASE.                                       |              |                        |         |
| DR | ProDom: PD000469; SOD_CU_ZN.3   |              |                        |         |
| KW | Copper; Oxidoreductase; Signal; 2inc.                                 |              |                        |         |
| FT | SIGNA1 1 20 POTENTIAL.  |              |                        |         |
| FT | CHAIN 21 517 PERIN.   |              |                        |         |
| SQ | SEQUENCE 517 AA: 57222 MW; 87B8FBFE85501E CRC64;                      |              |                        |         |
| QY | Query Match 87.3%; Score 434; DB 5; Length 517;                       |              |                        |         |
|    | Best Local Similarity 100.0%; Pred. No. 0;                            |              |                        |         |
|    | Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;          |              |                        |         |
|    | 64 GHGAVYLDELHVGFTSEDDHDDHGHGLHMLGDMSAGCDSIGELYNAPBEKHADPDGLG 123     |              |                        |         |

DB 84 GGGAGVLEHLVGNFSEDDHHHGLHLMGDSAGDSIGELYNAHPKHADPDGLG 143  
OY 124 DLVDDRGVNEVHYAMLDIDGTAPNTEALIGHSMTIIQSSHTDADTPASRIACVYGH 183  
DB 144 DLVDDRGVNEVHYAMLDIDGTAPNTEALIGHSMTIIQSSHTDADTPASRIACVYGH 203  
OY 184 GKARPEAAALHNELEDKTEHYAHCDVRSNTHOPKALHHNHGHTIDPKOYGYDLEVS 243  
DB 204 GKARPEAAALHNELEDKTEHYAHCDVRSNTHOPKALHHNHGHTIDPKOYGYDLEVS 263  
OY 244 HLEGFNVSDDHKLHVDQIYANGDLTSGCDNLGAKYDPHEHDYHSELGLDIDDDHGV 303  
DB 264 HLEGFNVSDDHKLHVDQIYANGDLTSGCDNLGAKYDPHEHDYHSELGLDIDDDHGV 323  
OY 304 VNESHRYSMINIFGDSVLGRSIAIHORHHLKSAKIAACVYGRGOSHPETVHRAKCYVR 363  
DB 324 VNESHRYSMINIFGDSVLGRSIAIHORHHLKSAKIAACVYGRGOSHPETVHRAKCYVR 383  
OY 364 PNTSTGLHHVSGSTFEQTPGSGTHMTADLKGFVSEDLSHHRYGOLHEMGDSHGC 423  
DB 384 PNTSTGLHHVSGSTFEQTPGSGTHMTADLKGFVSEDLSHHRYGOLHEMGDSHGC 443  
OY 424 HSLGRMTHGHDDADPKRPDGLDVIDDSHGIVHSTRFDHLNVEDLNARSLVIMOGSHE 483  
DB 444 HSLGRMTHGHDDADPKRPDGLDVIDDSHGIVHSTRFDHLNVEDLNARSLVIMOGSHE 503  
OY 484 VESERVACVYGRA 497  
DB 504 VESERVACVYGRA 517

## RESULT 2

AC P83148 PRELIMINARY; PRT: 20 AA.  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DE Major extracellular fluid protein (EP protein) (Fragment).  
OS Mytilus edulis (Blue mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
OC Mytiloidae; Mytilidae; Mytilus.  
NCBI\_TaxID=6550.

RC SEQUENCE, SUBUNIT, GLYCOSYLATION, AND MASS SPECTROMETRY.  
RX TISSUE-EXTRAPALLIAL FLUID:  
MEDLINE=21264951; PubMed=11084027;  
RA Harten S.J., Laue T.M., Chasteen N.D.;  
RT "Purification and characterization of a novel calcium-binding protein  
from the extrapallial fluid of the mussel, Mytilus edulis."  
RJ J. Biol. Chem. 276:4461-4468(2001).  
CC -1- FUNCTION: APPEARS TO BE A BUILDING BLOCK OF THE SOLUBLE ORGANIC  
MATRIX OF THE SHELL. THE PROTEIN BINDS CALCIUM.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- PTM: GLYCOSYLATED.  
CC -1- MASS SPECTROMETRY: MW=28340; METHOD=MALDI.  
CC -1- MISCELLANEOUS: HAS AN ISOELECTRIC POINT OF 4.43.  
KM Calcium-binding; Glycoprotein.  
FT NON\_TER 20  
SQ SEQUENCE 20 AA; 2351 MW; CBB0C9EDDE6F7451 CRC64;

Query Match 1.8%; Score 9; DB 5; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 DHHDDHND 20  
DB 4 DHHDDHND 12

## RESULT 3

OSHPNG

ID OSHPNG PRELIMINARY; PRT: 79 AA.  
AC OSHPNG; 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Vn91546h.  
GN Vn91546h.  
OS Halobacterium sp. (strain NRC-1).  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Halobacterium.  
NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20504483; PubMed=11016950;  
RA Ng W.V., Kennedy S.P., Mahlras G.G., Berquist B., Pan M.,  
RA Shukla H.D., Laake S.R., Bailga N.S., Thorsson V., Sirogna J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weli R., Goo Y.A.,  
RA Leitbauer B., Keller K., Cruz R., Danon M.J., Hough D.W.,  
RA Madocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
RA Isenberger T.A., Beck R.F., Pohlshöder M., Spudich J.L., Jung K.-H.,  
RA Alsm M., Freltas T., Hou S., Daniels C.J., Dennis P.P., Omar A.D.,  
RA Ehardt H., Love T.M., Liang P., Riley M., Hood L., Dassarina S.;  
RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
DR EMBL; AB050508; A619631.1; -;  
DR InterPro; IPR004045; GST\_Nterm.  
KW Complete proteome.  
SQ SEQUENCE 79 AA; 8511 MW; CAC5748CA97B4BE0 CRC64;

Query Match 1.8%; Score 9; DB 17; Length 79;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 LVDDRGV 133  
DB 55 LVDDRGV 63

## RESULT 4

ID OS5040 PRELIMINARY; PRT: 473 AA.  
AC OS5040; 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Ribulose biphosphate carboxylase large chain (Ec 4.1.1.39) (Rubisco  
large subunit).  
GN Rbcl OR CBRL.  
OS Thiohalophilus neapolitanus.  
OC Bacteria; Proteobacteria; gamma subdivision; Chromatiales;  
OC Halobacterellus.  
NCBI\_TaxID=927;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 23641;  
RX MEDLINE=96361901; PubMed=9696760;  
RA Baker S.H., Jin S., Aldrich H.C., Howard G.T., Shively J.M.;  
RT "Insertion mutation of the form I cbcl gene encoding ribulose  
biphosphate carboxylase/oxygenase (Rubisco) in Thiohalophilus  
neapolitanus results in expression of form II Rubisco, loss of  
RT carboxysomes, and an increased CO2 requirement for growth."  
RJ J. Bacteriol. 180:4133-4139(1998).

CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
ACTIVE SITE (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) - 3-  
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.  
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) - 2 3-  
CC PHOSPHO-D-GLYCERATE.  
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).

```

CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AF038430; AAC32549.1; -.
DR HSSP: P09657; 1BXN.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.N.1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW PHOTORESPIRATION.
SQ SEQUENCE 473 AA; 52635 MW; B84D2DE46CAF7D8 CRC64;

Query Match 1.8%; Score 9; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 NIFGDSVL 322
Db 385 NIFGDSVL 393

RESULT 5
OY 092B35 PRELIMINARY; PRT; 473 AA.
AC 092B35;
DT 01-MAY-1999 (TREMBLrel. 10; Created)
DT 01-MAY-1999 (TREMBLrel. 10; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20; Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
large subunit).
GN RCL.
OS Pseudomonas carboxydoflava.
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Hydrogenophaga.
OX NCBI_TaxID=47421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1083;
RA Kim I.M.; Lee S.N.;
RL Submitted (Apr.1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: U55037; AAD1031.1; -.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.N.1.
DR PROSITE: PS00157; RUBISCO_LARGE; 1.
DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW PHOTORESPIRATION.
SQ SEQUENCE 473 AA; 52692 MW; 87F82B96F8A35C59 CRC64;

Query Match 1.8%; Score 9; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 NIFGDSVL 322
Db 385 NIFGDSVL 393

RESULT 6
OY 086380 PRELIMINARY; PRT; 83 AA.
AC 086380;

```

```

DT 01-NOV-1998 (TREMBLrel. 08; Created)
DT 01-NOV-1998 (TREMBLrel. 08; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20; Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
large subunit) (Fragment).
GN RCL.
OS Nostoc commune.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=1178;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98317287; PubMed=9642201;
RA Rudi K.; Skulberg O.M.; Jakobsen K.S.;
RT "Evolution of cyanobacteria by exchange of genetic material among
phylogenetically related strains."
RL J. Bacteriol. 180:3453-3461(1998).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: Z94892; CAB08231.1; -.
DR InterPro: IPR000685; Rubisco_large.
DR Pfam: PF00016; Rubisco_large.N.1.
DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;
KW PHOTORESPIRATION.
FT NON_TER 1
SQ SEQUENCE 83 AA; 9013 MW; 00F555C8A51F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 IFGDSVL 322
Db 1 IFGDSVL 8

RESULT 7
OY 093A05 PRELIMINARY; PRT; 83 AA.
AC 093A05;
DT 01-DEC-2001 (TREMBLrel. 19; Created)
DT 01-DEC-2001 (TREMBLrel. 19; Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20; Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco
large subunit) (Fragment).
GN RCL.
OS Anabaena cf. lemmermannii 262.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=135961;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=262;
RA Gujger M.; Lyra C.; Henriksen P.; Coute A.; Humbert J.F.; Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
Anabaena and Aphanizomenon."
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.

```

CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 DR EMBL: AJ293142; CAC51565.1; -  
 DR InterPro: IPR000685; Rubisco\_large.  
 DR Pfam: PF00016; Rubisco\_large; 1.  
 KW Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;  
 KW Photorespiration.  
 RN NON\_TER 1  
 SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;  
 Query Match 1.6%; Score 8; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 315 IFGDDSVL 322  
 Db 1 IFGDDSVL 8  
 RESULT 8  
 ID Q93A03 PRELIMINARY; PRT; 83 AA.  
 AC Q93A03;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).  
 GN RBCL.  
 OS Aphanizomenon cf. gracile 271.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Aphanizomenon.  
 OX NCBI\_TaxID=136082;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=271;  
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;  
 RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon";  
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLICOLATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 DR EMBL: AJ293147; CAC51583.1; -  
 DR InterPro: IPR000685; Rubisco\_large.  
 DR Pfam: PF00016; Rubisco\_large; 1.  
 KW Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;  
 KW Photorespiration.  
 RN NON\_TER 1  
 SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;  
 Query Match 1.6%; Score 8; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 315 IFGDDSVL 322  
 Db 1 IFGDDSVL 8

AC Q93A00;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).  
 GN RBCL.  
 OS Anabaena cf. cylindrica 133.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 OX NCBI\_TaxID=136078;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=133;  
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;  
 RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon";  
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLICOLATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 DR EMBL: AJ293158; CAC51604.1; -  
 DR InterPro: IPR000685; Rubisco\_large.  
 DR Pfam: PF00016; Rubisco\_large; 1.  
 KW Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;  
 KW Photorespiration.  
 RN NON\_TER 1  
 SQ SEQUENCE 83 AA; 8990 MW; 906DEC461851F24D CRC64;  
 Query Match 1.6%; Score 8; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 315 IFGDDSVL 322  
 Db 1 IFGDDSVL 8  
 RESULT 10  
 ID Q939Z8 PRELIMINARY; PRT; 83 AA.  
 AC Q939Z8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO large subunit) (Fragment).  
 GN RBCL.  
 OS Anabaena crassa.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 OX NCBI\_TaxID=136074;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=215;  
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;  
 RT "Morphological and molecular comparison of the cyanobacterial genera Anabaena and Aphanizomenon";  
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-

```

CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293165; CAC51606.1; -.
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large.1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA: 9013 MW: 00F5555CA851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8

RESULT 11
Q93927 PRELIMINARY; PRT; 83 AA.
ID 093927;
AC 093927;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBCL.
OS Anabaena sp. PMC9701.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=136080;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9701;
RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
RT Anabaena and Aphanizomenon."
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293163; CAC51608.1; -.
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large.1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA: 9013 MW: 00F5555CA851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8

RESULT 12
Q93925 PRELIMINARY; PRT; 83 AA.
ID 093925;
AC 093925;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBCL.
OS Anabaena cf. cylindrica PMC9705.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
OX NCBI_TaxID=136079;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9705;
RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
RT Anabaena and Aphanizomenon."
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-
CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCULATE.
CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-
CC PHOSPHO-D-GLYCERATE.
CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
DR EMBL: AJ293165; CAC51610.1; -.
DR InterPro: IPR000685; RUBISCO_large.
DR Pfam: PF00016; RUBISCO_large.1.
KM Carbon dioxide fixation; lyase; Monooxygenase; Oxidoreductase;
KW Photorespiration.
FT NON_TER 1
SQ SEQUENCE 83 AA: 9013 MW: 00F5555CA851F25D CRC64;

Query Match 1.6%; Score 8; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 IFGDDSVL 322
DB 1 IFGDDSVL 8

RESULT 13
Q93924 PRELIMINARY; PRT; 83 AA.
ID 093924;
AC 093924;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (RUBISCO
DE large subunit) (Fragment).
GN RBCL.
OS Aphanizomenon cf. flos-aquae PMC9501.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Aphanizomenon.
OX NCBI_TaxID=136083;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PMC9501;
RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;
RT "Morphological and molecular comparison of the cyanobacterial genera
RT Anabaena and Aphanizomenon."
RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).
CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-
CC RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
CC ACTIVE SITE (BY SIMILARITY).

```

CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLICOLATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC EMBL: AJ293169; CAC51616.1; -  
 DR InterPro: IPR000685; Rubisco\_large.  
 DR Pfam: PF00016; Rubisco\_large.  
 DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;  
 KM Photorespiration.  
 FT NON\_TER 1  
 SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;  
 QY 315 IFGDDSVL 322  
 DB 1 IFGDDSVL 8

Query Match 1.6%; Score 8; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14  
 ID 093922 PRELIMINARY; PRT; 83 AA.  
 AC 093922;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco  
 large subunit) (Fragment).  
 GN RBCL.  
 OS Anabaena macrospora.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 OX NCBI\_TaxID=135103;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PMC9301;  
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;  
 RT "Morphological and molecular comparison of the cyanobacterial genera  
 Anabaena and Aphanizomenon";  
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLICOLATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC EMBL: AJ293161; CAC51631.1; -  
 DR InterPro: IPR000685; Rubisco\_large.  
 DR Pfam: PF00016; Rubisco\_large.  
 DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;  
 KM Photorespiration.  
 FT NON\_TER 1  
 SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;  
 QY 315 IFGDDSVL 322  
 DB 1 IFGDDSVL 8

Q939W9  
 ID Q939W9 PRELIMINARY; PRT; 83 AA.  
 AC Q939W9;  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Ribulose biphosphate carboxylase large chain (EC 4.1.1.39) (Rubisco  
 large subunit) (Fragment).  
 GN RBCL.  
 OS Anabaena menoltae.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 OX NCBI\_TaxID=136070;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=57;  
 RA Gugger M., Lyra C., Henriksen P., Couté A., Humbert J.F., Sivonen K.;  
 RT "Morphological and molecular comparison of the cyanobacterial genera  
 Anabaena and Aphanizomenon";  
 RL Int. J. Syst. Evol. Microbiol. 0:0-0(0).  
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-  
 RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC  
 CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF  
 THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH  
 REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME  
 ACTIVE SITE (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-  
 CC PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLICOLATE.  
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-  
 CC PHOSPHO-D-GLYCERATE.  
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.  
 CC EMBL: AJ293155; CAC51633.1; -  
 DR InterPro: IPR000685; Rubisco\_large.  
 DR Pfam: PF00016; Rubisco\_large.  
 DR Carbon dioxide fixation; Lyase; Monooxygenase; Oxidoreductase;  
 KM Photorespiration.  
 FT NON\_TER 1  
 SQ SEQUENCE 83 AA; 9013 MW; 00F5555CA851F25D CRC64;  
 QY 315 IFGDDSVL 322  
 DB 1 IFGDDSVL 8

Query Match 1.6%; Score 8; DB 2; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: November 29, 2002, 12:46:11  
 Job time : 36 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 10:10:43 ; Search time 70 seconds  
(without alignments)  
946.079 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 2794  
Sequence: 1 DGEQNDGKDKDHHDDHDD.....MOGHEVESERVACVIGRA 497

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq.101002.\*

1: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description         |
|------------|-------|-------------|--------|----------|---------------------|
| 1          | 2794  | 100.0       | 497    | AAV93750 | Amino acid sequence |
| 2          | 260   | 9.3         | 1529   | AAV97985 | CORC potassium cha  |
| 3          | 217   | 7.8         | 351    | AAV97985 | Sequence of Histid  |
| 4          | 173.5 | 6.2         | 244    | AAV97985 | Rat superoxide dis  |
| 5          | 173.5 | 6.2         | 244    | AAV97985 | Superoxide dismuta  |
| 6          | 172.5 | 6.2         | 221    | AAV97985 | GAG fusion protein  |
| 7          | 170   | 6.1         | 221    | AAV97985 | GAG fusion protein  |
| 8          | 169.5 | 6.1         | 210    | AAV97985 | GAG fusion protein  |
| 9          | 169.5 | 6.1         | 210    | AAV97985 | GAG fusion protein  |
| 10         | 169.5 | 6.1         | 228    | AAV97985 | GAG fusion protein  |

|    |       |     |      |    |          |                      |
|----|-------|-----|------|----|----------|----------------------|
| 11 | 168   | 6.0 | 205  | 13 | AAV97985 | GAG fusion protein   |
| 12 | 167   | 6.0 | 153  | 23 | AAV97985 | Metall-free apoprote |
| 13 | 167   | 6.0 | 154  | 23 | AAV97985 | Superoxide dismuta   |
| 14 | 167   | 6.0 | 189  | 13 | AAV97985 | GAG fusion protein   |
| 15 | 167   | 6.0 | 205  | 13 | AAV97985 | Novel human diapo    |
| 16 | 167   | 6.0 | 598  | 22 | AAV97985 | Novel human diapo    |
| 17 | 166.5 | 6.0 | 1939 | 22 | AAV97985 | Fungal stress respo  |
| 18 | 165.5 | 5.9 | 153  | 23 | AAV97985 | GAG fusion protein   |
| 19 | 165.5 | 5.9 | 186  | 13 | AAV97985 | GAG fusion protein   |
| 20 | 165.5 | 5.9 | 199  | 13 | AAV97985 | GAG fusion protein   |
| 21 | 165.5 | 5.9 | 212  | 13 | AAV97985 | GAG fusion protein   |
| 22 | 165.5 | 5.9 | 392  | 21 | AAV97985 | HIV-1 env4/human s   |
| 23 | 164.5 | 5.9 | 153  | 23 | AAV97985 | Superoxide dismuta   |
| 24 | 164.5 | 5.9 | 392  | 21 | AAV97985 | SOD/env-4 fusion p   |
| 25 | 164   | 5.9 | 151  | 23 | AAV97985 | Superoxide dismuta   |
| 26 | 164   | 5.9 | 152  | 23 | AAV97985 | zeta mays protein f  |
| 27 | 163.5 | 5.9 | 153  | 23 | AAV97985 | Rice superoxide d    |
| 28 | 163   | 5.8 | 152  | 23 | AAV97985 | Paddy copper zinc    |
| 29 | 163   | 5.8 | 152  | 23 | AAV97985 | GAG fusion protein   |
| 30 | 163   | 5.8 | 183  | 13 | AAV97985 | GAG fusion protein   |
| 31 | 163   | 5.8 | 183  | 13 | AAV97985 | GAG fusion protein   |
| 32 | 163   | 5.8 | 184  | 13 | AAV97985 | GAG fusion protein   |
| 33 | 163   | 5.8 | 185  | 13 | AAV97985 | GAG fusion protein   |
| 34 | 163   | 5.8 | 186  | 13 | AAV97985 | GAG fusion protein   |
| 35 | 163   | 5.8 | 189  | 13 | AAV97985 | GAG fusion protein   |
| 36 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 37 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 38 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 39 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 40 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 41 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 42 | 163   | 5.8 | 192  | 13 | AAV97985 | GAG fusion protein   |
| 43 | 163   | 5.8 | 203  | 13 | AAV97985 | GAG fusion protein   |
| 44 | 163   | 5.8 | 203  | 13 | AAV97985 | GAG fusion protein   |
| 45 | 163   | 5.8 | 206  | 13 | AAV97985 | GAG fusion protein   |

#### ALIGNMENTS

|  |  |
|--|--|
| RESULT 1   |  |
| AAV93750   | standard; Protein: 497 AA.   |
| AAV93750   |  |
| 03-OCT-2000  | (first entry)  |
| Amino acid sequence of a serine protease inhibitor protein.    |  |
| Serine protease inhibitor; green-lipped mussel; anti-chrombin; |  |
| divalent metal cation binding activity; dietary supplement;    |  |
| anticoagulant.   |  |
| OS   | Perna canaliculus.   |
| XX   |  |
| PN   | WO200039165-A1.  |
| XX   |  |
| PD   | 06-JUL-2000.   |
| XX   |  |
| PF   | 23-DEC-1999; 99WO-NZ00227.   |
| XX   |  |
| PR   | 23-DEC-1998; 98NZ-0333568.   |
| XX   |  |
| PA   | 23-JUL-1999; 99NZ-0336906.   |
| XX   |  |
| PI   | (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.                  |
| XX   |  |
| DR   | Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;                   |
| XX   |  |
| DR   | WPI: 2000-452375/39  |
| XX   |  |
| DR   | N-PSDB: AAA47150, AAA47151.  |
| XX   |  |
| PI   | New Perna canaliculus serine protease inhibitor protein exhibiting |

|                           |   |
|---------------------------|---|
| Pt                        | anti-thymbin activity and divalent metal cation binding activity,       |
| Pt                        | useful as an anticoagulant agent and as a dietary supplement -          |
| Xx                        |   |
| Ps                        | Claim 2; Page 2; 44pp: English.   |
| Xx                        |   |
| Cc                        | The present sequence represents a serine protease inhibitor             |
| Cc                        | protein. The protein is isolated from the green-lipped mussel           |
| Cc                        | (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity   |
| Cc                        | and divalent metal cation binding activity. The serine protease         |
| Cc                        | inhibitor protein has a molecular weight of about 55 kDa Daltons,       |
| Cc                        | the protein, and its fragments, are useful in medicaments, in food,     |
| Cc                        | as dietary supplements or as biomedication agents. In the dietary       |
| Cc                        | supplements, the protein is associated with or bound to at least one    |
| Cc                        | divalent cation (such as calcium, magnesium or zinc) of dietary         |
| Cc                        | significance. The proteins or their fragments are also useful as        |
| Cc                        | anticoagulant agents.   |
| Xx                        |   |
| Sq                        | Sequence 497 AA:  |
|                           |   |
| Query Match               | 100.0%; Score 2794; DB 21; Length 497;                                  |
| Best Local Similarity     | 100.0%; Pred. No. 4,6e-254;   |
| Matches 497; Conservative | 0; Mismatches 0; Indels 0; Gaps   |
| Oy                        | 1 DGECCNDGKKRDDHDDHDDHHDDDDDDDEMTYAAOCSEMERPNMMASSLHHNHGSITEL 60        |
| Dd                        | 1 DGECNDSGGKRRDHDHDDHDDHHDDDDDDDETMYAOCSEMERPNMMASSLHHNHGSITEL 60       |
| Oy                        | 61 SQRKGAVUETLELVGFNTSEDHDDHNGSLNLMIGMSACSDSIGELYUAANEKRNADRG 120       |
| Dd                        | 61 SQKHGAUVLETLVGFNTSEDHDDHNGSLNLMIGMSACSDSIGELYUAANEKRNADRG 120        |
| Oy                        | 121 DLGRVUDRDRGVNVENNYAMLDLGSTARPTFALIGNSHTLLQSGSHTDATPRASACSY 180      |
| Dd                        | 121 DLGRVUDRDRGVNVENNYAMLDLGSTARPTFALIGNSHTLLQSGSHTDATPRASACSY 180      |
| Oy                        | 181 IGSKARPTAALLNHELEDKTENTYAHNCDAVSNTPOKALNNHNCGSTIDFKOVGYGDLE 240     |
| Dd                        | 181 IGHKARPETAALLNHELEDKTENTYAHNCDAVSNTPOKALNNHNCGTIDFKOVGYGDLE 240     |
| Oy                        | 241 VSYHLGEPNVSDDKDKHLDVOUYANGDLTGSCDMYGAKYRDNEHYNSELDAGDIHDDO 300      |
| Dd                        | 241 VSYLHEFPNVSDDKDKHLDVOUYANGDLTGSCDMYGAKYRPNDYHSELGLDGIHDDO 300       |
| Oy                        | 301 HGVAUNESHRSWININGEGDSDVLGRSAIAYHQBNILHKSAKICCVYIGGOSHPRETVNRACK 360 |
| Dd                        | 301 HGVAUNESHRSWININGEGDSDVLGRSAIAYIQORNLHKSAKICCVYIGGOSHPRETVNRACK 360 |
| Oy                        | 361 VVRPNTESDTLIHHNVGSITFEOTPGSGSTNMTPADLKGFNVSEDSLHNRHGVDLHWGMDS 420   |
| Dd                        | 361 VVRPNTESDTLIHHNVGSITFEOTPGSGSTMTPADLKGFNVSEDSLHNRHGVDLHWGMDS 420    |
| Oy                        | 421 HGHSLSLRMVNGHDDANDPKRPRLDGADVDDSHGIVHSTFRFDLANVEDLNASTLYVMOG 480    |
| Dd                        | 421 HGHSLSLRMVNGHDDANDPKRPRLDGADVDDSHGIVHSTRFDLANVEDLNASTLYVMOG 480     |
| Oy                        | 481 GHEYESERVACSYIGRA 497   |
| Dd                        | 481 GHEYESERVACSYIGRA 497   |
| RESULT 2                  |   |
| AAR97985                  |   |
| ID                        | AAR97985 standard; Protein: 1529 AA.                                    |
| XX                        |   |
| AC                        | AAR97985:   |
| XX                        |   |
| DT                        | 26-OCT-1996 (first entry)   |
| XX                        |   |
| DE                        | CORK potassium channel protein alternative sequence.                    |
| XX                        |   |
| KX                        | CORK: potassium channel; nematode; pore-forming domain;                 |
| KV                        | transmembrane helix; N-glycosylation site; potassium-agonist;           |
| KW                        | potassium-antagonist; drug screening; nematocidae; anthelmintic;        |

|                       |   |  |     |
|-----------------------|---|--|-----|
| KW                    |   | cardiac disorder.  |     |
| XX                    | OS  | <i>Caenorhabditis elegans</i> .  |     |
| XX                    |   |  |     |
| FM                    |   | W09613520-A1.  |     |
| XX                    |   |  |     |
| FD                    |   | 09-MAY-1996.   |     |
| XX                    |   |  |     |
| PF                    |   | 25-OCT-1995; 95WO-US14364.   |     |
| XX                    |   |  |     |
| PR                    |   | 31-OCT-1994; 94US-0332312.   |     |
| XX                    |   |  |     |
| PA                    |   | (AMCY ) AMERICAN CYANAMID CO.  |     |
| P1                    |   | Pausch MH, Price LA;   |     |
| DR                    |   | WPI: 1996-239450/24.   |     |
| XX                    |   |  |     |
| PT                    |   | Potassium channel genes from <i>Drosophila melanogaster</i> and            |     |
| XX                    |   | <i>Caenorhabditis elegans</i> - useful in assaying substances to determine |     |
| PT                    |   | effects on cell growth, and in inhibiting nematode and insect pests        |     |
| XX                    |   |  |     |
| PS                    |   | Claim 13; Page 49-53; 79pp; English.                                       |     |
| XX                    |   |  |     |
| CC                    |   | This sequence is presented as a potassium channel sequence encoded by      |     |
| CC                    |   | the CORK gene from <i>Caenorhabditis elegans</i> . An alternative sequence |     |
| CC                    |   | is given in AAR92315. The protein may be expressed in a heterologous       |     |
| CC                    |   | host cell to assay substances to determine effects on cell growth.         |     |
| CC                    |   | Potassium-channelists or potassium-antagonists identified by this method   |     |
| CC                    |   | may be used as nematocides, anthelmintics or in therapy of cardiac         |     |
| CC                    |   | disorders, etc.  |     |
| XX                    |   |  |     |
| S0                    |   | Sequence 1529 AA:  |     |
| Query Match           | 9.38%; Score 260; DB 17; Length 1529;                                 |  |     |
| Best Local Similarity | 19.4%; Pred. No. 6,6e-13;   |  |     |
| Matches               | 96; Conservative 62; Mismatches 190; Indels 148; Gaps                 |  | 25. |
| OY                    | 6 NDGSGKKDDNHH--HDDNHDDHD--DDETMVAQSEMERPMASLLNNHVGSTEL               | 60   |     |
| DB                    | 180 HNVAVNDVBDVVANVNNHNDNDNDNDNDNDNDNDNDNDNDNDNDNDVDVAN-DDAANDNDIVANN | 238  |     |
| OY                    | 61 SQKNGAGUUDELTLVGFNTSDDH-DH-NIGLILMLGLMSAGCISGEIYMANREKH--          | 116  |     |
| OY                    | 239 AVNANH-----DVBVNDADNHNHNDH-----DAANNHNDHNDH                       | 175  |     |
| DB                    | 117 --ADPGLSLVDDDRGVUNEVCHNYAMLDDTARTETALIGNMTILLGSHTDADTPAS          | 174  |     |
| DB                    | 276 AAABVDANDH--DVNDH-----DADAVIDDNA-----NVBDAADA                     | 311  |     |
| OY                    | 175 RIACSVIGNHAKARETAALLNHELE-----DKTEN-----ANCDSVRSYTNOKPAL          | 221  |     |
| DB                    | 312 -----HNVADDDADVNNNDVANDVANNNDVBAVNVBAADVBAADVANNNAHD              | 363  |     |
| OY                    | 222 NNNHNGTIDRKQUGUDLEVSYNLEGSFVNDKGNKLIDUVTUANSGLTGSCCNLGAKDU        | 281  |     |
| DB                    | 364 NNNHNAVBNNDHVNIVANNHNN-----DVNDHBBVVDIAVDINABVD-----VDNHNND       | 414  |     |
| OY                    | 282 RHED--YNELEDLDLDHDD-----DH-----GVNESIRYSWIMIFGGDDVLSRSTALND       | 330  |     |
| DB                    | 415 VHVDDDNHNVNNHBAVDANBDNBAVDNBAVDANBANNNHNVANBVNDHNA-----NH         | 467  |     |
| OY                    | 331 RDILNKSAKTASCUTIGRGSGHREIYNRAKCVVRRTTESTGLNNVSGSTIFFEQPGSGTH      | 390  |     |
| DB                    | 468 DNNHNDADADVBAVDNALDNNHNBVA-----DHBVNDHNN-----H                    | 505  |     |
| OY                    | 391 MTADLKGFNVSEDLSHNRKGVOLNEMGSKN--GCHSLGRMYNSHNDADRKPGRDLGDV        | 448  |     |
| DB                    | 506 VNHD-----DNAABVNDHDAANDVDVAVNVBAHVAANNHNAADVAAHD-----             | 549  |     |
| OY                    | 449 IDDSHGIVNSTRTFDH  | 464  |     |
| DB                    | 550 -DDEVNANDHDAABH   | 564  |     |

| Result | Accession   | Standard                             | Protein          | Length |
|--------|---|--------------------------------------|------------------|--------|
| 3      | AA024393  | standard                             | Protein; 351 AA. | 351    |
| AC     | AA024393;   |                                      |                  |        |
| DT     | 22-NOV-1992   | (first entry)                        |                  |        |
| DE     | Sequence of Histidine-rich protein (HsRP).                          |                                      |                  |        |
| XX     |   |                                      |                  |        |
| XX     | Malaria vaccine; Histidine-rich protein; cytoadherence.             |                                      |                  |        |
| OS     | Plasmodium lophurae.  |                                      |                  |        |
| FT     | Key   | Location/Qualifiers                  |                  |        |
| FT     | Modified-site   | 40..42                               |                  |        |
| FT     | Peptide   | /label= potential glycosylation site |                  |        |
| FT     | Peptide   | 1..23                                |                  |        |
| FT     | Peptide   | /label= signal                       |                  |        |
| FT     | Peptide   | 24..47                               |                  |        |
| FT     | Peptide   | /label= pro-peptide                  |                  |        |
| XX     | US5116965-A.  |                                      |                  |        |
| XX     | 25-MAY-1992.  |                                      |                  |        |
| XX     | 26-AUG-1986;  | 86US-0900401.                        |                  |        |
| XX     | 26-AUG-1986;  | 86US-0900401.                        |                  |        |
| XX     | (SLOK ) SLOAN KETTERING INST CANCER.                                |                                      |                  |        |
| PA     | Pologe L, Ravetch JV;   |                                      |                  |        |
| XX     | WPI: 1992-199590/24.  |                                      |                  |        |
| DR     | N-PSDB; AA024393.   |                                      |                  |        |
| XX     |   |                                      |                  |        |
| PT     | Histidine-rich protein associated with Plasmodium knob phenotype -  |                                      |                  |        |
| PT     | and DNA encoding it, used for in vitro diagnosis of P.              |                                      |                  |        |
| XX     | Falciparum infection.   |                                      |                  |        |
| XX     | Disclosure; Fig 7A-B; 29pp; English.                                |                                      |                  |        |
| XX     |   |                                      |                  |        |
| CC     | Two variants of HsRP are produced by P. falciparum. One is          |                                      |                  |        |
| CC     | associated with what is referred to as "knobby phenotype" (K30) and |                                      |                  |        |
| CC     | "knobless phenotype" (K-). The "knobby" and "knobless" phenotypes   |                                      |                  |        |
| CC     | have been implicated in cytoadherence, which is characteristic of   |                                      |                  |        |
| CC     | erythrocyte infection. It has now been found that cDNA expressing   |                                      |                  |        |
| CC     | both K+ and K- HsRP can be obtained by the use of P. lophurae HsRP  |                                      |                  |        |
| CC     | expressing DNA. The genomic clone (AA025532) is encoded in two      |                                      |                  |        |
| CC     | exons, separating the signal peptide-encoding sequence from the     |                                      |                  |        |
| CC     | pro-sequence, confirming that synthesis of the protein occurs via   |                                      |                  |        |
| CC     | the preproprotein. Oligo. probes synthesised to the signal          |                                      |                  |        |
| CC     | peptide-encoding exon reveal multiple homologous DNA sequences in   |                                      |                  |        |
| CC     | the P. lophurae genome. The sequence of mature proteins is arranged |                                      |                  |        |
| CC     | in numerous tandem repeats with up to nine histidine residues in a  |                                      |                  |        |
| CC     | row, similar to other Plasmodium proteins for which sequence data   |                                      |                  |        |
| CC     | have so far been reported.  |                                      |                  |        |
| XX     |   |                                      |                  |        |
| XX     | Sequence 351 AA:  |                                      |                  |        |
| XX     |   |                                      |                  |        |
| XX     | Query Match 7.8%; Score 217; DB 13; Length 351;                     |                                      |                  |        |
| XX     | Best Local Similarity 19.7%; Pred. No. 8.6e-12;                     |                                      |                  |        |
| XX     | Matches 84; Conservative 29; Mismatches 104; Indels 150; Gaps       |                                      |                  |        |
| XX     | 12 DDNDNDNDNDNDNDNDDEDTNHYAO-CAMERNPMASLLNHYNGSTLSQKNGCAVLT 70      |                                      |                  |        |
| XX     | 67 EENERRNERRNERRNERRNERRNERRNERRNERRNERRNERRNERRNERRNERRN 120      |                                      |                  |        |
| XX     | 71 ELNLRVETSEDND 130      |                                      |                  |        |

|      |  |     |   |   |                      |
|------|--|-----|---|---|----------------------|
| D6   |  | 121 | NNNDAANNNEEENNNNAADNN-----  | -NNEEJNN--  | 152                  |
| OY   |  | 131 | GUVNEVNHAKMLDOSTARPELALGSGMPTLLQSSPHDAOTRASRTACSVGNKAREP  | 190   |                      |
|      |  |     | :   :   |   |                      |
| D6   |  | 153 | --AAANHNHWF-----  | ---NNINIGUNNNNARNINNINNIAFNN-----                 | NNNNARN- 195         |
| OY   |  | 191 | AALNLLEDEEDTGENAUCSDVRKYSTPNORXALNNHVUGSTDIFQGVUSGLSEVSUYLEGRNV   | 250   |                      |
|      |  |     | :   :   | :   :   |                      |
| D6   |  | 196 | -----NNNNNNNARNNNNN-----  | ---NNIARNNNNNNNNNNNNNNGSNNNNNNNNNNNNNN-           | 245                  |
| OY   |  | 251 | SDDIKCHLIIDUYGTANGDLTGSCSNLSAKVDREDFUNSELGSLODIIDDYGAVNESNYR  | 310   |                      |
|      |  |     | :   :   | :   :   |                      |
| D6   |  | 246 | --NNNNHHNRDANN-----   | ---NNINDANN--                                     | ---NNNDANNNNNNNN 281 |
| OY   |  | 311 | SWNIIEODDSVLASRLATIJORDYLKSKAKTLASCYSVGKSQSPRELYNQRAKCVVRRPYESTG  | 370   |                      |
|      |  |     | :   :   |   |                      |
| D6   |  | 282 | -----NNANDNN--  | ---NNINDAN-----                                   | 295                  |
| OY   |  | 371 | LNNHSGSTTFEGDTREGSTMHTADIKGFNVSEDLSNNRGSQLVEEMQMSKCSISLRMT  | 430   |                      |
|      |  |     | :   :   | :   :   |                      |
| D6   |  | 296 | -NNH-----   | ---NNNNRNANNNNNNNNNNNNNN-NNNDANNNNNNNNNDANNNN 337 |                      |
| OY   |  | 431 | HGHDDAH 437   |   |                      |
|      |  |     | :   :   |   |                      |
| D6   |  | 338 | NNNNDAH 344   |   |                      |
| <br> |  |     |   |   |                      |
|      | RESULT 4   |     |   |   |                      |
| ID   | AARe67409  |     |   |   |                      |
| AC   | AARe67409 standard; Protein; 244 AA.   |     |   |   |                      |
| XX   | AARe67409;   |     |   |   |                      |
| DT   |  |     |   |   |                      |
| XX   | 08-AUG-1995 (first entry)  |     |   |   |                      |
| DE   | Rat superoxide-dismutase-B.  |     |   |   |                      |
| XX   |  |     |   |   |                      |
| KW   | Superoxide-dismutase; EC-SOD-B; hyperoxia; emphysema;<br>respiratory distress syndrome; retinopathy; inflammation;<br>autoimmune disease.  |     |   |   |                      |
| OS   | Rattus sp.   |     |   |   |                      |
| FH   |  |     |   |   |                      |
| FH   | Key  |     | Location/Qualifiers   |   |                      |
| FT   | Peptide  |     | 1..20   |   |                      |
| FT   |  |     | /Label= Sig_peptide   |   |                      |
| FT   | Peptide  |     | 21..244   |   |                      |
| FT   |  |     | /label= Mat_peptide   |   |                      |
| FT   |  |     | /note= "preferred polypeptides span mature peptide<br>amino acids 199-213, 1-55, 53-72, 71-89,<br>87-102, 100-108, 119-127, 133-141, 138-156,<br>157-166, 173-181, 181-189, 195-214 and<br>219-224 (claim 9, page 33-34)" |   |                      |
| FT   |  |     |   |   |                      |
| XX   | EPR27486-A.  |     |   |   |                      |
| PD   |  |     |   |   |                      |
| XX   | 07-DEC-1994.   |     |   |   |                      |
| Pf   |  |     |   |   |                      |
| XX   | 04-JUN-1993:   |     | 93EP-0401438.   |   |                      |
| XX   |  |     |   |   |                      |
| PR   | 04-JUN-1993:   |     | 93EP-0401438.   |   |                      |
| XX   |  |     |   |   |                      |
| PA   | (INNO-) INNOGENERICS NV SA.  |     |   |   |                      |
| XX   |  |     |   |   |                      |
| PI   | Bettadapura J, Slegers H, Willems J, Zwijsen A;  |     |   |   |                      |
| DR   |  |     |   |   |                      |
| XX   | WPt: 1995-008070/02.   |     |   |   |                      |
| DR   | N-PSDB; AAQ75336.  |     |   |   |                      |
| Pf   |  |     |   |   |                      |
| XX   | Purified extracellular superoxide dismutase B form - useful e.g.<br>for treating lung tissue damage, radiation damage, retinopathy in<br>infants, inflammatory and auto-immune disease |     |   |   |                      |
| XX   |  |     |   |   |                      |



CC for targeting them to specific cell surfaces or substrates. The  
 CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
 CC protein with SOD to increase stability, plasma half-life and ease  
 CC of purification of SOD. SOD is useful for reduction of tissue damage  
 CC caused by oxygen radicals and is used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid and osteo arthritis.  
 CC See also AAR24225-35, AAR27932-51.

XX Sequence 221 AA:

Query Match 6.28; Score 172.5; DB 13; Length 221;  
 Best Local Similarity 23.58; Pred. No. 6.8e-08;  
 Matches 60; Conservative 40; Mismatches 72; Indels 83; Gaps 10;

QY 225 VHGTFDFKO-VGYGDELYSYHLEGFNVSDDKDLHDVOIYANGDGLTSGCDNLGAKYDP- 282  
 DB 15 VQGIINFEGKESNGPVKWKWSIKGLT-----EGLHGFVHNEFGDMTAGCTSGAPHFNPL 68  
 QY 283 -----HEDYHSELGLDLIDHDDHGVNESHRYSWINIFGDDSVLGRSIAIHQR-DHLH 335  
 DB 69 SRKHGGRPKDERHNVGLDGNVTADKGVADVSIEDSVISLGSDDHSIIIGRTLVYHEKADDLG 128  
 QY 336 K-----SAKIACVYIGRQGS-----HP-----EIVHRAKCVVRFNTESTG 370  
 DB 129 KGNBESTKTGNAGSRILACGVIGIAGCGHHRHREMKKRVEDLMH----- 175  
 QY 371 LHHNVSGSTFEQTPGSGTHTADLKGFNVSEDLSHRHGVOLHEMGDMSHGCHSLGRMY 430  
 DB 176 -----HHPREMK-KRVEDLMHHRHREMK-----RVEDLM 205  
 QY 431 HGHDADHDKRPDGL 445  
 DB 206 HRHHPREMKKRVEDL 220

RESULT 7

ID AAR24232 standard; Protein: 221 AA.

XX AAR24232;

DT 25-NOV-1992 (first entry)

DE GAG fusion protein with SOD according to a formula.

XX Glycosamino:glycan; superoxidisedismutase; tissue damage;  
 KW autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..49 /note= "GAG binding motif"

FT Peptide 50..221 /note= "SOD"

XX WO9207935-A.

XX 14-MAY-1992.

XX 01-NOV-1991; 91WO-US08105.

XX 01-NOV-1990; 90US-0608539.

XX 02-NOV-1990; 90US-0608569.

XX (SCRI ) SCRIPPS RES INST.

PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;  
 PI Mullenbach JC, Parge HE, Tainer JA;

XX WPI: 1992-183671/22.

XX Fusion proteins with glycosamino:glycan-binding and

PT superoxidisedismutase activities - reduce tissue damage caused by  
 PT superoxide radicals; useful in treating autoimmune diseases e.g.  
 PT rheumatoid arthritis and osteoarthritis  
 XX Claim 8; Fig 1: 140pp; English.

CC The fusion protein comprising the a glycosaminoglycan binding region  
 CC and human superoxide dismutase, joined via a linker region was  
 CC constructed according to the formula M-(Z-M)3-SOD where Z is the peptide  
 CC -HRHHPREMKKRVEDL-. The fusion protein is useful for extending  
 CC in vivo lifetimes of biologically active cps. such as SOD and  
 CC for targeting them to specific cell surfaces or substrates. The  
 CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
 CC protein with SOD to increase stability, plasma half-life and ease  
 CC of purification of SOD. SOD is useful for reduction of tissue damage  
 CC caused by oxygen radicals and is used in the treatment of autoimmune  
 CC diseases e.g. rheumatoid and osteo arthritis.  
 CC See also AAR24225-35, AAR27933-51.

XX Sequence 221 AA:

Query Match 6.18; Score 170; DB 13; Length 221;  
 Best Local Similarity 25.78; Pred. No. 1.2e-07;  
 Matches 55; Conservative 34; Mismatches 67; Indels 58; Gaps 9;

QY 185 KAPETAAALHHELE-EDKTEHYAHCDVRSNTWOPKALHH----- 224  
 DB 10 KRVEDLMHRRHHPREMKKRVEDLMH-----RHHPREMKKRVEDLMATKAVAVLKGDGP 63  
 QY 225 VHGTFDFKO-VGYGDELYSYHLEGFNVSDDKDLHDVOIYANGDGLTSGCDNLGAKYDP- 282  
 DB 64 VQGIINFEGKESNGPVKWKWSIKGLT-----EGLHGFVHNEFGDMTAGCTSGAPHFNPL 117  
 QY 283 -----HEDYHSELGLDLIDHDDHGVNESHRYSWINIFGDDSVLGRSIAIHQR-DHLH 335  
 DB 118 SRKHGGRPKDERHNVGLDGNVTADKGVADVSIEDSVISLGSDDHSIIIGRTLVYHEKADDLG 177  
 QY 336 K-----SAKIACVYIGRQGS-----HP 352  
 DB 178 KGNBESTKTGNAGSRILACGVIGIAGCGHHRHP 211

RESULT 8

ID AAR24231 standard; Protein: 192 AA.

XX AAR24231;

DT 25-NOV-1992 (first entry)

DE GAG fusion protein with SOD according to a formula.

XX Glycosamino:glycan; superoxidisedismutase; tissue damage;  
 KW autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..172 /note= "SOD"

FT Peptide 173..192 /note= "GAG binding motif"

XX WO9207935-A.

XX 14-MAY-1992.

XX 01-NOV-1991; 91WO-US08105.

XX 01-NOV-1990; 90US-0608539.

XX 02-NOV-1990; 90US-0608569.

XX (SCRI ) SCRIPPS RES INST.

```
XX Bolstein M, Fisher C, Griffin JH, Hallelwell RA, Kuhn L;  
PI Mullenbachgt, Parge HE, Tainer JA;  
XX WPI: 1992-183671/22.  
DR  
XX Fusion proteins with glycosamino:glycan-binding and  
PT superoxidismitase activities - reduce tissue damage caused by  
PT superoxide radicals' useful in treating autoimmune diseases e.g.  
PT rheumatoid arthritis and osteoarthritis  
XX  
PS Claim 7; Fig 1; 140pp; English.  
CC The fusion protein comprising the a glycosaminoglycan binding region  
CC and human superoxide dismutase, joined via a linker region was  
CC constructed according to the formula SOD-(M-2)-M where 2 is the peptide  
CC -RPRSGSKRRKRLKRS-. The fusion protein is useful for extending  
CC the in vivo lifetimes of biologically active epds, such as SOD and  
CC for targeting them to specific cell surfaces or substrates. The  
CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
CC of purification of SOD. SOD is useful for reduction of tissue damage  
CC caused by oxygen radicals and is used in the treatment of autoimmune  
CC diseases e.g. rheumatoid and osteo-arthritis.  
CC See also AAR24225-35, AAR27932-51.  
XX  
SO Sequence 192 AA;  
Query Match 6.1%; Score 169.5; DB 13; Length 192;  
Best Local Similarity 29.2%; Pred. No. 1.1e-07;  
Matches 50; Conservative 34; Mismatches 54; Indels 33; Gaps 7;  
OY 225 VHGTFIDFKQ-VGYGDLEVSYLEGFNVSDDKDHLHDVQIYANGDLTSGCDNLGAKYDP-282  
DB 15 VQGIINEQKESNGPVKVMGSIKGLT-----EGDGHGFVHGFBDNAGCTSAQPHFNPL 68  
OY 283 -----HEDYHSELGDLIDDDHGVNESHRYSMINIFGDDSVLGRSIAIHOR-DHLH 335  
DB 69 SRKHGKPRDEERHVGDLGVNTADKGVADVSIEDSVISLSDHSHIIGRTLVVHKKADLDG 128  
OY 336 K-----SAKIACCVIGRGS-----HP-EIYHRAKCVVRPTES 368  
DB 129 KGNNESTKTGNAGSRACGVIGIAGPGRHRRHPRMKKRVEDLMRVPRRES 179  
RESULT 9  
AAR24230  
ID AAR24230 standard; Protein: 210 AA.  
XX  
AC AAR24230;  
XX  
DT 25-NOV-1992 (first entry)  
XX  
DE GAG fusion protein with SOD according to a formula.  
XX  
KW Glycosamino:glycan: superoxidismitase; tissue damage;  
XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..172  
FT /note= "SOD"  
FT 173..210  
FT Peptide /note= "GAG binding motif"  
XX  
XX W09207935-A.  
XX  
XX 14-MAY-1992.  
XX  
XX 01-NOV-1991; 91MO-US08105.  
XX  
XX 01-NOV-1990; 90US-0608539.  
XX
```

```
PR 02-NOV-1990; 90US-0608569.  
XX  
XX (SCRI.) SCRIIPS RES INST.  
XX  
XX Bolstein M, Fisher C, Griffin JH, Hallelwell RA, Kuhn L;  
PI Mullenbachgt, Parge HE, Tainer JA;  
XX WPI: 1992-183671/22.  
DR  
XX Fusion proteins with glycosamino:glycan-binding and  
PT superoxidismitase activities - reduce tissue damage caused by  
PT superoxide radicals' useful in treating autoimmune diseases e.g.  
PT rheumatoid arthritis and osteoarthritis  
XX  
PS Claim 7; Fig 1; 140pp; English.  
CC The fusion protein comprising the a glycosaminoglycan binding region  
CC and human superoxide dismutase, joined via a linker region was  
CC constructed according to the formula SOD-(M-2)-M where 2 is the peptide  
CC -RPRSGSKRRKRLKRS-. The fusion protein is useful for extending  
CC the in vivo lifetimes of biologically active epds, such as SOD and  
CC for targeting them to specific cell surfaces or substrates. The  
CC glycosaminoglycan (GAG) binding protein is formed into a fusion  
CC of purification of SOD. SOD is useful for reduction of tissue damage  
CC caused by oxygen radicals and is used in the treatment of autoimmune  
CC diseases e.g. rheumatoid and osteo-arthritis.  
CC See also AAR24225-35, AAR27932-51.  
XX  
SO Sequence 210 AA;  
Query Match 6.1%; Score 169.5; DB 13; Length 210;  
Best Local Similarity 29.2%; Pred. No. 1.2e-07;  
Matches 50; Conservative 34; Mismatches 54; Indels 33; Gaps 7;  
OY 225 VHGTFIDFKQ-VGYGDLEVSYLEGFNVSDDKDHLHDVQIYANGDLTSGCDNLGAKYDP-282  
DB 15 VQGIINEQKESNGPVKVMGSIKGLT-----EGDGHGFVHGFBDNAGCTSAQPHFNPL 68  
OY 283 -----HEDYHSELGDLIDDDHGVNESHRYSMINIFGDDSVLGRSIAIHOR-DHLH 335  
DB 69 SRKHGKPRDEERHVGDLGVNTADKGVADVSIEDSVISLSDHSHIIGRTLVVHKKADLDG 128  
OY 336 K-----SAKIACCVIGRGS-----HP-EIYHRAKCVVRPTES 368  
DB 129 KGNNESTKTGNAGSRACGVIGIAGPGRHRRHPRMKKRVEDLMRVPRRES 179  
RESULT 10  
AAR24227  
ID AAR24227 standard; Protein: 228 AA.  
XX  
AC AAR24227;  
XX  
DT 25-NOV-1992 (first entry)  
XX  
DE GAG fusion protein with SOD according to a formula.  
XX  
KW Glycosamino:glycan: superoxidismitase; tissue damage;  
XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..172  
FT /note= "SOD"  
FT 173..228  
FT Peptide /note= "GAG binding motif"  
XX  
XX W09207935-A.  
XX  
XX 14-MAY-1992.  
XX
```

```

PF 01-NOV-1991: 91WO-US08105.
XX
XX 01-NOV-1990: 90US-0608539.
PR 02-NOV-1990: 90US-0608569.
XX
PA (SCRI ) SCRIPPS RES INST.
PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;
PI Mullenbachgt, Parge HE, Talner JA;
XX WPI: 1992-183671/22.
XX
PT Fusion proteins with glycosamino:glycan-binding and
PT superoxidismutase activities - reduce tissue damage caused by
PT super:oxide radicals, useful in treating autoimmune diseases e.g.
PT rheumatoid arthritis and osteoarthritis
XX
PS Claim 7: Fig 1: 140pp; English.
XX
XX The fusion protein comprising the a glycosaminoglycan binding region
CC and human superoxide dismutase, joined via a linker region was
CC constructed according to the formula SOD-(M-Z)-M where Z is the peptide
CC RNPESGKRRKRRKLPs. The fusion protein is useful for extending
CC the in vivo lifetimes of biologically active cpds. such as SOD and
CC for targeting them to specific cell surfaces or substrates. The
CC glycosaminoglycan (GAG) binding protein is formed into a fusion
CC protein with SOD to increase stability, plasma half-life and ease
CC of purification of SOD. SOD is useful for reduction of tissue damage
CC caused by oxygen radicals and is used in the treatment of autoimmune
CC diseases e.g. Rheumatoid and osteo-arthritis.
CC See also AAR24225-35, AAR27932-51.
XX
SQ Sequence 228 AA:
Query Match 6.1%; Score 169.5; DB 13; Length 228;
Best Local Similarity 29.2%; Pred. No. 1.4e-07;
Matches 50; Conservative 34; Mismatches 54; Indels 33; Gaps 7;
OY 225 VHGTFPKO-VGYGDELYSHLEGFNVSDDKHLDHVDQIYANDLTSGCDNIGAKYDP- 282
DB 15 VQGIINFEQKESGPKVKGSIKGLT-----EGLHGPHVHEFGDNTAGCTSGAPHNPL 68
OY 283 -----HEDYHSELGLDGIHDDHGCVNESHRSWIMIFGDDSVLGRSAIAHQ-R-DHIA 335
DB 69 SRKHGPKDEERHVDLGWVTDKDGADVSTEDSVISLSGDHSIIIGRTLVHEKADDLG 128
OY 336 K-----SAKIACCVTGRGQS-----HP-ETVHRACKVVRPNTESG 368
DB 129 KGNNESTKTGNAGSRLACGVIGIAOGPGHRRHPRMKRVEDLMRVPRES 179
RESULT 11
AAR27935
ID AAR27935 standard; Protein: 205 AA.
AC AAR27935;
XX
XX 25-NOV-1992 (first entry)
XX
XX GAG fusion protein with SOD according to a formula.
XX
XX glycosamino:glycan; superoxidismutase; tissue damage;
XX autoimmune disease; rheumatoid arthritis; osteoarthritis; ss.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Peptide 1..172 "SOD"
XX FT /note="SOD"
XX FT 173..205
XX FT Peptide /note="GAG binding motif"
XX
XX WQ0207935-A.

```

```

XX
XX 14-MAY-1992.
PD
XX
XX 01-NOV-1991: 91WO-US08105.
PF
XX
XX 01-NOV-1990: 90US-0608539.
PR 02-NOV-1990: 90US-0608569.
XX
PA (SCRI ) SCRIPPS RES INST.
PI Boissinot M, Fisher C, Griffin JH, Hallewell RA, Kuhn L;
PI Mullenbachgt, Parge HE, Talner JA;
XX WPI: 1992-183671/22.
XX
XX Fusion proteins with glycosamino:glycan-binding and
PT superoxidismutase activities - reduce tissue damage caused by
PT super:oxide radicals, useful in treating autoimmune diseases e.g.
PT rheumatoid arthritis and osteoarthritis
XX
PS Claim 8: Fig 1: 140pp; English.
XX
XX The fusion protein comprising the a glycosaminoglycan binding region
CC and human superoxide dismutase, joined via a linker region was
CC constructed according to the formula SOD-(M-Z)-M where Z is the peptide
CC -RHHPREKRRKVEDL-. The fusion protein is useful for extending
CC the in vivo lifetimes of biologically active cpds. such as SOD and
CC for targeting them to specific cell surfaces or substrates. The
CC glycosaminoglycan (GAG) binding protein is formed into a fusion
CC protein with SOD to increase stability, plasma half-life and ease
CC of purification of SOD. SOD is useful for reduction of tissue damage
CC caused by oxygen radicals and is used in the treatment of autoimmune
CC diseases e.g. Rheumatoid and osteo-arthritis.
CC See also AAR24225-35, AAR27932-51.
XX
SQ Sequence 205 AA:
Query Match 6.0%; Score 168; DB 13; Length 205;
Best Local Similarity 24.7%; Pred. No. 1.6e-07;
Matches 54; Conservative 35; Mismatches 56; Indels 74; Gaps 9;
OY 225 VHGTFPKO-VGYGDELYSHLEGFNVSDDKHLDHVDQIYANDLTSGCDNIGAKYDP- 282
DB 15 VQGIINFEQKESGPKVKGSIKGLT-----EGLHGPHVHEFGDNTAGCTSGAPHNPL 68
OY 283 -----HEDYHSELGLDGIHDDHGCVNESHRSWIMIFGDDSVLGRSAIAHQ-R-DHIA 335
DB 69 SRKHGPKDEERHVDLGWVTDKDGADVSTEDSVISLSGDHSIIIGRTLVHEKADDLG 128
OY 336 K-----SAKIACCVTGRGQS-----HP-----ETVHRACKVVRPNTESG 370
DB 129 KGNNESTKTGNAGSRLACGVIGIAOGPGHRRHPRMKRVEDLMHR----- 175
OY 371 LHHVSGSITFEQTPGSGTHMTADLKGFNVSEDLSHNRH 409
DB 176 -----HHPREMK--KRVEDLMHRRH 193
RESULT 12
AAP70979
ID AAP70979 standard; Protein: 153 AA.
AC AAP70979;
XX
XX 03-MAY-1991 (first entry)
XX
XX Metalfree apoprotein for metalloprotein-albumin conjugate (II).
XX
XX Metalloprotein; apoprotein; albumin; antibody; antlynosin.
XX
XX Key Location/Qualifiers
XX Disulfide-bond 57..146
XX

```

|    |   |
|----|---|
| PN | DE3718476-A.  |
| PP | 03-DEC-1987.  |
| XX |   |
| XX | 02-JUN-1987; 87DE-3718476.  |
| XX |   |
| XX | 02-JUN-1986; 86GB-0013315.  |
| XX |   |
| XX | (DEBI- ) DEBIOPHARM SA.   |
| XX |   |
| PI | Degenhohl R;  |
| XX |   |
| XX | WPI, 1987-343120/49.  |
| DR |   |
| XX |   |
| PT | Biologically active metallo:protein albumin conjugates - produced       |
| PT | by crosslinking a metal-free:apoprotein in the presence of              |
| PT | albumin and adding metal to the purified conjugate                      |
| XX |   |
| PS | Claim 3; Page 1; 5pp; German.   |
| XX |   |
| XX |   |
| CC | To Al may be attached H or acetyl and the protein is obtainable by      |
| CC | the procedure of US4340675.   |
| CC | The procedure is crosslinked in the presence of human albumin and opt.  |
| CC | an antibody fragment. The metal-free conjugate is purified, and then    |
| CC | biological activity is restored by adding the desired metal, eg         |
| CC | copper-zinc, to the conjugate. The opt. antibody is anti-lysozyme       |
| CC | material. The product enables the metalloprotein to be directed         |
| CC | to a particular tissue or organ for therapeutic purposes, the           |
| CC | immunogenicity and antigenicity to be reduced, and/or the               |
| CC | bioavailability and therapeutic activity to be increased.               |
| CC | The MP-albumin-anti-lysozyme conjugates are esp. suitable for directing |
| CC | MP to myosin, a protein of heart muscle.                                |
| CC | See also AAP70978.  |
| XX |   |
| XX |   |
| XX | Sequence 153 AA;  |

|                       |  |  |       |                                |
|-----------------------|--|--|-------|--------------------------------|
| Query Match           | 6.0%;  | Score 167;   | DB 8; | Length 153;                    |
| Best Local Similarity | 30.1%;   | Pred. No. 1.3e-07;   |       |                                |
| Matches               | 43;  | Conservative   | 26;   | Mismatches 48; Indels 26; Gaps |
| OY                    | 225  | VHGIDPKQVGYCD--LEVSYHLEGFNVSDDKDHLHDQIVANGDLTSGCDNIGAKYDPH   | 283   |                                |
| Db                    | 13   | VSGVGVKFEQASESEPTTVSYETLANSPPNAREGFHHE-----GDATNGCVSAGPHNPFP | 67    |                                |
| OY                    | 284  | EDVYS-----PLGSLGDIHDDGHVYNSHRYSWIMIPDGSDVLRSTAIHQ--RPHLH     | 335   |                                |
| Db                    | 68   | KKTHGAPTDEVRHVGDMGVNKTDEKENVAKGSPKSDLILILPTSVYGRSVYIAHAGDDLG | 127   |                                |
| OY                    | 336  | K-----SAKIACCVTG   | 346   |                                |
| Db                    | 128  | KGDTESLKTGTGNAGPRPACGVIG                                     | 150   |                                |
| RESULT 13             |  |  |       |                                |
| AAAS2499              |  |  |       |                                |
| ID                    | AAAS2499   | standard; Protein; 154                                       | AA.   |                                |
| AC                    | AAAS2499;  |  |       |                                |
| XX                    |  |  |       |                                |
| DT                    | 03-JUL-2002  | (first entry)  |       |                                |
| XX                    |  |  |       |                                |
| DE                    | Superoxide dismutase protein #28.                                |  |       |                                |
| XX                    |  |  |       |                                |
| KW                    | Selenoprotein; HIV; Ebola virus; cancer; immune system disorder. |  |       |                                |
| XX                    |  |  |       |                                |
| OS                    | Saccharomycetes cerevisiae.                                      |  |       |                                |
| XX                    |  |  |       |                                |
| PN                    | US6303295-B1.  |  |       |                                |
| XX                    |  |  |       |                                |
| PD                    | 16-OCT-2001.   |  |       |                                |
| XX                    |  |  |       |                                |
| FF                    | 12-JUL-1996;   | 96US-0679493.  |       |                                |
| XX                    |  |  |       |                                |

|    |   |               |
|----|---|---------------|
| PR | 14-JUL-1995;  | 95US-001203P. |
| PR | 01-SEP-1995;  | 95US-003112P. |
| XX |   |               |
| PA | (UYGB-) UNIV GEORGIA RES FOUND INC.                                       |               |
| XX |   |               |
| PI | Taylor EW, Nadimpalli RG, Ramanathan CS;                                  |               |
| XX |   |               |
| DR | WPI: 2002-024734/03.  |               |
| XX |   |               |
| PT | New selenoprotein for use in detecting certain viruses, e.g. human        |               |
| PT | immunodeficiency virus (HIV) or Ebola, cancer and immune system           |               |
| PT | disorders -   |               |
| XX |   |               |
| PS | Disclosure: Columns 77-88; 140pp; English.                                |               |
| XX |   |               |
| CC | The present invention relates to selenoproteins encoded in the genome of  |               |
| CC | a virus, where the coding sequence of the selenoprotein is genetically    |               |
| CC | engineered for expression in a nucleic acid construct. The invention also |               |
| CC | discloses a method for identifying selenoprotein coding sequences, for    |               |
| CC | detecting certain viruses (e.g. HIV or Ebola), cancer and immune system   |               |
| CC | disorders. The present sequence was used to illustrate the invention.     |               |
| XX |   |               |
| XX | Sequence  | 154 AA;       |

|                       |       |   |        |                                   |
|-----------------------|-------|---|--------|-----------------------------------|
| Query Match           | 30.1% | Score 167;  | DB 23; | Length 154;                       |
| Best Local Similarity | 3.1%  | Pred. No. 1.3e-07;  |        |                                   |
| Matches               | 43;   | Conservative  | 26;    | Mismatches 48; Indels 26; Gaps 5; |
| QY                    | 225   | VHGTIDFKQGVGD-LEVSYLEGNSVDHDKHIDVQIYANDLISGCDNIGAKAYDP    | 283    |                                   |
| Db                    | 14    | VSGVYVFEGQASEEPTTYSYELAGNSPNAENGFIHNE-----GDALNGCVSAGPHNF | 68     |                                   |
| QY                    | 284   | EDVHS-----ELGGLDHHDDHGVNNSHRYSMIIEGDSVGLRSTAIHQ-RDHL      | 335    |                                   |
| Db                    | 69    | KRTHGAPTDVRYHMGVYKKTDENGVAKSGFSDSLILQPTSVYGRSVYIHAGDDLG   | 128    |                                   |
| QY                    | 336   | K-----SAKIACVYG   | 346    |                                   |
| Db                    | 129   | KGPTESLKTGNAGPRPACGVIG                                    | 151    |                                   |

|          |   |
|----------|---|
| RESULT   | 14  |
| AAR27936 |   |
| ID       | AAR27936 standard; Protein; 189 AA.   |
| XX       |   |
| AC       | AAR27936;   |
| XX       |   |
| DI       | 25-NOV-1992 (first entry)   |
| XX       |   |
| DE       | GAG fusion protein with SOD according to a formula.   |
| XX       |   |
| KW       | Glycosamino:glycan; superoxidisedismutase; tissue damage;<br>autoimmune disease; rheumatoid arthritis; osteoarthritis; ss |
| XX       |   |
| OS       | Synthetic.  |
| XX       |   |
| FH       | Key   |
| FT       | Peptide   |
| FT       | /note= "SOD"  |
| FT       | 173..189  |
| FT       | /note= "GAG binding motif"  |
| XX       |   |
| PN       | M09207935-A.  |
| XX       |   |
| PD       | 14-MAY-1992.  |
| XX       |   |
| PF       | 01-NOV-1991; 91MO-US08105.  |
| XX       |   |
| PR       | 01-NOV-1990; 90US-0608539.  |
| PR       | 02-NOV-1990; 90US-0608569.  |
| XX       |   |
| PA       | (SCRI ) SCRIPPS RES INST.   |
| XX       |   |





THIS PAGE BLANK (USPTO)

101,000,000

THE

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:32:35 ; Search time 26 Seconds  
(without alignments)  
562.430 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 2794  
Sequence: 1 DGECCNDGKNDKHHDDHD.....MOGHEVSERVACVIGRA 497

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                             |
|------------|-------|-------------|--------|-------|---|
| 1          | 173.5 | 6.2         | 244    | 4     | US-08-679-493A-188<br>Sequence 188, App |
| 2          | 168   | 6.0         | 150    | 2     | US-08-722-050-9<br>Sequence 9, Appl     |
| 3          | 167   | 6.0         | 154    | 4     | US-08-679-493A-211<br>Sequence 211, App |
| 4          | 166   | 5.9         | 151    | 2     | US-08-722-050-10<br>Sequence 10, Appl   |
| 5          | 165.5 | 5.9         | 152    | 2     | US-08-722-050-12<br>Sequence 12, Appl   |
| 6          | 164.5 | 5.9         | 153    | 4     | US-08-679-493A-207<br>Sequence 191, App |
| 7          | 164   | 5.9         | 151    | 4     | US-08-679-493A-191<br>Sequence 201, App |
| 8          | 163.5 | 5.9         | 153    | 4     | US-08-679-493A-201<br>Sequence 202, App |
| 9          | 161.5 | 5.8         | 153    | 4     | US-08-679-493A-202<br>Sequence 202, App |
| 10         | 160.5 | 5.7         | 152    | 6     | 5171680-3<br>Sequence 209, App          |
| 11         | 159   | 5.7         | 166    | 4     | US-08-679-493A-209<br>Sequence 192, App |
| 12         | 158   | 5.7         | 151    | 4     | US-08-679-493A-192<br>Sequence 203, App |
| 13         | 158   | 5.7         | 152    | 4     | US-08-679-493A-203<br>Sequence 5, Appl  |
| 14         | 157.5 | 5.6         | 618    | 1     | US-08-668-381A-5<br>Sequence 5, Appl    |
| 15         | 156.5 | 5.6         | 154    | 6     | 5290690-5<br>Sequence 12, Appl          |
| 16         | 156.5 | 5.6         | 1021   | 1     | US-07-910-760-12<br>Sequence 12, Appl   |
| 17         | 156.5 | 5.6         | 1021   | 1     | US-08-440-519-12<br>Sequence 12, Appl   |
| 18         | 156.5 | 5.6         | 1021   | 1     | US-08-440-519-12<br>Sequence 12, Appl   |
| 19         | 155.5 | 5.6         | 144    | 6     | 5252476-5<br>Sequence 12, Appl          |
| 20         | 155.5 | 5.6         | 152    | 6     | 5196335-1<br>Sequence 204, App          |
| 21         | 155.5 | 5.6         | 153    | 4     | US-08-679-493A-204<br>Sequence 204, App |
| 22         | 155.5 | 5.6         | 153    | 4     | 5290690-6<br>Sequence 4, Appl           |
| 23         | 155.5 | 5.6         | 154    | 4     | US-09-126-109-4<br>Sequence 7, Appl     |
| 24         | 153.5 | 5.5         | 152    | 2     | US-08-722-050-7<br>Sequence 2, Appl     |
| 25         | 152.5 | 5.5         | 152    | 2     | US-09-291-562-2<br>Sequence 3, Appl     |
| 26         | 152.5 | 5.5         | 153    | 2     | US-08-722-050-3<br>Sequence 86, Appl    |
| 27         | 152   | 5.4         | 841    | 1     | US-08-350-884-86<br>Sequence 86, Appl   |

|    |       |     |     |   |   |
|----|-------|-----|-----|---|---|
| 28 | 152   | 5.4 | 841 | 1 | US-08-709-173-86<br>Sequence 86, Appl   |
| 29 | 152   | 5.4 | 841 | 2 | US-08-709-177-86<br>Sequence 190, App   |
| 30 | 151   | 5.4 | 201 | 4 | US-08-679-493A-190<br>Sequence 8, Appl  |
| 31 | 150   | 5.4 | 151 | 2 | US-08-722-050-8<br>Sequence 208, App    |
| 32 | 150   | 5.4 | 152 | 4 | US-08-679-493A-208<br>Sequence 196, App |
| 33 | 150   | 5.4 | 151 | 4 | US-08-679-493A-196<br>Sequence 194, App |
| 34 | 148   | 5.3 | 222 | 6 | 5472691-3<br>Sequence 2, Appl           |
| 35 | 148   | 5.3 | 222 | 6 | 5472691-3<br>Sequence 43, Appl          |
| 36 | 148   | 5.3 | 240 | 2 | US-08-023-980B-45<br>Sequence 53, Appl  |
| 37 | 148   | 5.3 | 240 | 2 | US-08-486-953A-53<br>Sequence 186, App  |
| 38 | 148   | 5.3 | 240 | 4 | US-08-679-493A-186<br>Sequence 186, App |
| 39 | 148   | 5.3 | 240 | 4 | US-08-679-493A-186<br>Sequence 186, App |
| 40 | 148   | 5.3 | 240 | 6 | 5472691-2<br>Sequence 186, App          |
| 41 | 147.5 | 5.3 | 309 | 1 | US-08-161-406-2<br>Sequence 186, App    |
| 42 | 147   | 5.2 | 151 | 2 | US-08-722-050-3<br>Sequence 197, App    |
| 43 | 146.5 | 5.2 | 202 | 4 | US-08-679-493A-197<br>Sequence 197, App |
| 44 | 145   | 5.2 | 151 | 4 | US-08-679-493A-153<br>Sequence 197, App |
| 45 | 145   | 5.2 | 152 | 4 | US-08-679-493A-206<br>Sequence 206, App |

## ALIGNMENTS

```
RESULT 1
US-08-679-493A-188
; Sequence 188, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: TAYLOR, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 188
; LENGTH: 244
; TYPE: PRT
; ORGANISM: rat
US-08-679-493A-188

Query Match      6.2%  Score 173.5; DB 4; Length 244;
Best Local Similarity 29.7%; Pred. No. 1e-08;
Matches 54; Conservative 28; Mismatches 61; Indels 39; Gaps 10;

QY 194 LHHLEDEKTEHYACDVRSNTQPKAL---HHHVGITDFKQVGYCD-LEYSYHLEGF 248
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 55 LGRREADAREHNAVCRY-----OPSAIPLPDPOITGLVLFQIOLGSSRLLEASFLEGF 109

QY 249 NVSDDKHDLHVOIYANGDLTSGCDNLGAKYDPHEHYHSELGLDLDIHDDHG--VYNE 306
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 110 PABQNTSNA--IHVHEFGDLSGCESTGPHYNP-----LCVPHQHPGDFGNFYVD 160

QY 307 ----SHREYEW--INIGDSSVYLGSAIHO--RDHLKSA-----KIACCVIGRG 348
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 GRMKRMKLATSLAGPSHILGRVAVYHAGEDDLGKGNASVQNGNARRLACCVIGRS 220

QY 349 QS 350
      |
Db 221 NS 222

RESULT 2
US-08-722-050-9
; Sequence 9, Application US/08722050
; Patent No. 5871729
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: ROSEN, CRAIG A.
```

```

;
; APPLICANT: FRASER, CLAIRE M.
; APPLICANT: GOCAYNE, JEANNINE D.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,050
; FILING DATE: 23-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,757
; FILING DATE: 11-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1020001/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-722-050-9

Query Match
Best Local Similarity 31.9%; Score 168; DB 2; Length 150;
Matches 46; Conservative 22; Mismatches 46; Indels 30; Gaps 5;

OY 225 VHGTFDKQVGYDLEVSYLEGFNVSDDKHLDVQIYANGDLTSCDNLGAKYDP-- 282
DB 12 VKGIFFSQEGDPTTYTGSISL-----KGLGFFVHALSDTTNGCMSTGPHNPFV 65
OY 283 -----HEDYSELGDLGDHDDHGVNESHRYSWINIFGDSVLSGRSTAIHQ-RDHL 334
DB 66 KKHGAPTEVRRVGMGVNKTDENGVAKGSPKDSLILKIGPTSVGRSVYIHAGDDDLG 128
OY 336 K-----SAKIACCVIG 346
DB 124 GKGHELSKSTGNAGRGVACGII 147

RESULT 3
US-08-679-493A-211
; Sequence 211, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 211
```

```

;
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-08-679-493A-211

Query Match
Best Local Similarity 30.1%; Score 167; DB 4; Length 154;
Matches 43; Conservative 26; Mismatches 48; Indels 26; Gaps 5;

OY 225 VHGTFDKQVGYDLEVSYLEGFNVSDDKHLDVQIYANGDLTSCDNLGAKYDPH 283
DB 14 VSGVYKFEQASESEPTTYSYEIAGNSPNAERGFTHIER-----GDATNGCVSAGPHNPF 68
OY 284 EDYH-----ELGDLGDIHDDHGVNESHRYSWINIFGDSVLSGRSTAIHQ-RDHL 335
DB 69 KKHGAPTEVRRVGMGVNKTDENGVAKGSPKDSLILKIGPTSVGRSVYIHAGDDDLG 128
OY 336 K-----SAKIACCVIG 346
DB 129 KGTESLKTGNAGRPACGII 151

RESULT 4
US-08-722-050-10
; Sequence 10, Application US/08722050
; Patent No. 5871729
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: ROSEN, CRAIG A.
; APPLICANT: FRASER, CLAIRE M.
; APPLICANT: GOCAYNE, JEANNINE D.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,050
; FILING DATE: 23-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,757
; FILING DATE: 11-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1020001/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-722-050-10

Query Match
Best Local Similarity 30.3%; Score 166; DB 2; Length 151;
Matches 43; Conservative 24; Mismatches 49; Indels 26; Gaps 5;

OY 225 VHGTFDKQVGYDLEVSYLEGFNVSDDKHLDVQIYANGDLTSCDNLGAKYDPH 284
```

```
Db 13 VQGTIFHEQKAGCEPWLSGQITGL-TEGQHGFFHXY-----GDNTOGCTAGPHFNPHS 66
QY 285 DYHS-----ELGDLGDIHDDHGCVNESHRYSMINIFGDDSVLGRSTAIHQ-RDHLK 336
Db 67 KKHGPADEHRYVGLGNVTAGKDVANVSTIEDRVISLSEHSIIGRTVYVHEKODDLCK 126
QY 337 -----SAKIACCVIG 346
Db 127 GGNESTKTKGNAGSRILACGVIG 148

RESULT 5
US-08-722-050-12
; Sequence 12, Application US/08722050
; Patent No. 5871729
; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: ROSEN, CRAIG A.
; APPLICANT: PRASER, CLAIRE M.
; APPLICANT: GOCAYNE, JENNIFER D.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,050
; FILING DATE: 23-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/725,757
; FILING DATE: 11-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488,1020001/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 152 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-722-050-12

Query Match 5.9%; Score 165.5; DB 2; Length 152;
Best Local Similarity 30.8%; Pred. NO. 3e-08;
Matches 44; Conservative 25; Mismatches 47; Indels 27; Gaps 6;
```

```
RESULT 6
US-08-679-493A-207
; Sequence 207, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 207
; LENGTH: 153
; TYPE: PRT
; ORGANISM: horse
US-08-679-493A-207

Query Match 5.9%; Score 164.5; DB 4; Length 153;
Best Local Similarity 31.0%; Pred. NO. 3.7e-08;
Matches 45; Conservative 26; Mismatches 43; Indels 31; Gaps 6;

QY 225 VHGTFDKQVGYGDLVSYHLEGF---NVSDDKRDLHDVOIYANGDLTSGCDNLGAKYDP 281
Db 14 VHGVIHFEOQEGGPHV---LKGFEGTLKGDHGFVHERF-----GDNTOGCTAGAHFN 65
QY 282 P-----HEDYSELGDLGDIHDDHGCVNESHRYSMINIFGDDSVLGRSTAIHQ-RDHLK 333
Db 66 PLKKGKGGKDEHRYVGLGNVTADENKADVDKDSVLSKHSIIGRTVYVHEKOD 125
QY 334 LHK-----SAKIACCVIG 346
Db 126 LKGNESTKTKGNAGSRILACGVIG 150

RESULT 7
US-08-679-493A-191
; Sequence 191, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 191
; LENGTH: 151
; TYPE: PRT
; ORGANISM: corn
US-08-679-493A-191

Query Match 5.9%; Score 164; DB 4; Length 151;
Best Local Similarity 31.7%; Pred. NO. 4.1e-08;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;
```

Mon Dec 2 12:20:01 2002

us-09-868-760-7.rai

Page 4

Db 67 KEHGAEDEENRHAGDGLGNVTAGADGVANINWDSQ<sub>1</sub>PLTGPSNIIIGRAVVVHADPDLGK 126  
 QY 337 -----SAKIACVYG 346  
 Db 127 GGHLSKSTGNGAGRVACGIIIG 148  
 : : : : :  
 : : : : :

```

US-08-679-493A-201
US-Result 8
? Sequence 201, Application US/08679493A
? Patent No. 6303295
? GENERAL INFORMATION:
? APPLICANT: TAYLOR, Ethan W.
? TITLE/ABSTRACT: SELENOPROTEINS, CODING SEQUENCES AND METHODS
? CURRENT APPLICATION NUMBER: US/08/679,493A
? CURRENT FILING DATE: 1996-07-11
? PRIOR APPLICATION NUMBER: 60/001203
? PRIOR FILING DATE: 1995-07-14
? PRIOR APPLICATION NUMBER: 60/003,112
? PRIOR FILING DATE: 1995-09-01
? NUMBER OF SEQ ID NOS: 216
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 201
? LENGTH: 153
? TYPE: PRT
? ORGANISM: Mus sp.
US-08-679-493A-201

```

|    | Query Match   | Score | DB        | Length             |
|----|---|-------|-----------|--------------------|
|    | Similarity  | Pred. | No.       | Mismatches         |
| Oy | 225 VHGIRFKQVGYCD-LEVSYLHEGFNSDDHDLHDVQIYANDLTLSCGDNLCARQDP   | 5.9%  | 163..5    | 153;               |
| Dd | 14 VQGIIHEKQASEPVLVLSQTIGL-TSEQGHFNHQ-----GDNQGTSAQPHNHP      | 30.1% | 4..6e-08; | Indels 27; Gaps 6; |
| Oy | 284 EDVYS-----ELGGIDIHDDHGVNESRKYRWIMIFGDSVLRSLAIHQ-RDHL      |       |           | 335                |
| Dd | 68 SKKHGCADEERHVGDIAGNTAGKDQVANVTEDRVYISLGESHITIGRTVYHEKKQDDL |       |           | 127                |
| Oy | 336 K-----SAKIACCVTGRQ  |       |           | 349                |
| Dd | 128 KGNESTKTGNAGSRLLACGVIGIAQ                                 |       |           | 153                |

```

RESULT 9
US-08-679-493A-202
Sequence 202, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: TAYLOR, Ethna W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
FILE REFERENCE: 15-35 NUMBER: US/08/679,493A
CURRENT FILING DATE: 1996-07-11
PRIORITY FILING DATE: 1995-07-14
PRIORITY APPLICATION NUMBER: 60/001203
PRIORITY FILING DATE: 1995-07-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 202
LENGTH: 153
TYPE: PRT
ORGANISM: rat
US-08-679-493A-202

```

| Qy                    | 225              | VHGTFDEKQGVGD-LEVSHTLEGFNVSDDHDLHDDQIYANGDLTSGCDNLGAKYDPH | 283                |
|-----------------------|------------------|---|--------------------|
| Query Match           | 5.8%             | Score 161.5; DB 4   | Length 153;        |
| Best Local Similarity | 30.1%            | Pred. No. 7.2e-08;  |                    |
| Matches 44;           | Conservative 24; | Mismatches 51;  | Indels 27; Gaps 6; |

Db 14 VQCVIHFEOKAGSGEVPVVSQITQL-TEEGHGFIHVOY-----GNTGGCTTAPRHNPH 67

QY 284 EDYH-----ELDLDDIDHDDHGVVNSHRYSMINIFGDSVLGRSIAIHOR-DHLH 335

Db 68 SKKHGPADEKHNHVDLCLNVAAGKGVANNSIEDRVISLGEHSIIIGRTMVYHEKODLIG 127

QY 336 K-----SAKTIACVYIGRQ 349

Db 128 KGNENSTRTGNAGSLKCGVIGIAQ 153

```

RESULT 10
5171680-3
PATENT NO. 5171680
APPLICANT: MULLENBACH, GUY T.; HALLEWELL, ROBERT A.; VALEZUELA,
PABLO
TITLE OF INVENTION: SUPEROXIDE DISMUTASE ANALOGS HAVING NOVEL
BINDING PROPERTIES
NUMBER OF SEQUENCES: 15
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/561,442
FILING DATE: 01-AUG-1990
SEQ ID NO: 3
LENGTH: 152
5171680-3

```

[illegible]

```

US-08-679-493A-209
US-08-679-493A-209
Sequence 209, Application US/08679493A
Patent No. 6303295
GENERAL INFORMATION:
APPLICANT: Taylor, Ethan W.
TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
SUBSEQUENT PRIORITY NUMBER: US/08/679,493A
CURRENT FILING DATE: 1996-07-12
PRIOR APPLICATION NUMBER: 60/001203
PRIOR FILING DATE: 1995-07-14
PRIOR APPLICATION NUMBER: 60/003,112
PRIOR FILING DATE: 1995-09-01
NUMBER OF SEQ ID NOS: 216
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 209
LENGTH: 166
TYPE: PRT
ORGANISM: loggerhead
US-08-679-493A-209

```

```

Query Match      5.7%   Score 159; DB 4;   Length 166;
Best Local Similarity 28.9%   Pred. NO. 1.4e-07;
Matches 41;   Conservative 29;   Mismatches 46;   Indels 26;   Gaps 5
Oy  225 VHGTFDEKQVGVDDELYSHLEGFVNSDDHDKHLMDVIVYANGULYSGCGNGLAKAKDP-- 282
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  27 VKGIIVIEQGNQGNPVLISGLITGL-TEGKHGFIHVEF-----GNTNGCTSAGAHNPPQ 80

```

|  |                          |   |                   |
|--|--------------------------|---|-------------------|
| Oy   | 283                      | - - - HEDYHSELGLDGIHHDDHGYYNESHRSYSWINFGDDSVLGRSLATIHOR-DHLHK   | 336               |
| Dd   | 81                       | KNHGQGDNERHVVDLGIVANKSGVAEVCIKDSLILSTLSGSQSIIIGRMVYHNKREDDLGR   | 140               |
| Oy   | 337                      | -----SAKIACCVIS   | 346               |
| Dd   | 141                      | GGNDESKLTGGNAGSRKLACGVVG  | 162               |
| RESULT 12  |                          |   |                   |
| US-08-679-493A-192   |                          |   |                   |
| Sequence 192,  | Application US/08679493A |   |                   |
| Patent No.   | 6303295                  |   |                   |
| GENERAL INFORMATION:   |                          |   |                   |
| APLICANT:  | Taylor, Ethan W.         |   |                   |
| TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS |                          |   |                   |
| FILE REFERENCE:  | 55-95                    |   |                   |
| CURRENT APPLICATION NUMBER:                                      | US/08/679,493A           |   |                   |
| PRIOR FILING DATE:   | 1996-07-12               |   |                   |
| PRIOR APPLICATION NUMBER:  | 60/001203                |   |                   |
| PRIOR FILING DATE:   | 1995-07-14               |   |                   |
| PRIOR APPLICATION NUMBER:  | 60/003,112               |   |                   |
| PRIOR FILING DATE:   | 1995-09-01               |   |                   |
| NUMBER OF SEQ ID NOS:  | 216                      |   |                   |
| SOFTWARE:  | PatentIn Ver. 2.0        |   |                   |
| SEQ ID NO  | 192                      |   |                   |
| LENGTH:  | 151                      |   |                   |
| TYPE:  | PRF                      |   |                   |
| ORGANISM:  | rice                     |   |                   |
| US-08-679-493A-192   |                          |   |                   |
| Query Match  |                          |   |                   |
| Best Local Similarity  | 5.7%                     | Score 158;  | DB 4; Length 151; |
| Matches 44;  | Conservative 21;         | Mismatches 51;  | Indels 26; Gaps   |
| Oy   | 225                      | VHGTFPKOYGVDLEVSYHLEGFNVSDDKHDKHLDPDYLYANGDLTSCGDNLGAKPDP--     | 282z              |
| Dd   | 13                       | VKGITFFSQEGDGPPHSVGSVSI-----KPLGHGFHYHALGDPTTGCMOSTGPHPNPYG     | 66                |
| Oy   | 283                      | ----HEDYHSELGLDGIHHDDHGYYNESHRSYSWINFGDDSVLGRSLATIH-QRDHLRK     | 336z              |
| Dd   | 67                       | KENGAPQDENRRHAAGDLGNTGTAGCAGVANNVSDSQLPVTGANSHIIGRAVVYHADPDLLGR | 120z              |
| Oy   | 337                      | -----SAKIACCVIS   | 346               |
| Dd   | 127                      | GGHELKTGTGNAGRACGCILIS  | 148               |
| RESULT 13  |                          |   |                   |
| US-08-679-493A-203   |                          |   |                   |
| Sequence 203,  | Application US/08679493A |   |                   |
| Patent No.   | 6303295                  |   |                   |
| GENERAL INFORMATION:   |                          |   |                   |
| APLICANT:  | Taylor, Ethan W.         |   |                   |
| TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS |                          |   |                   |
| FILE REFERENCE:  | 55-95                    |   |                   |
| CURRENT APPLICATION NUMBER:                                      | US/08/679,493A           |   |                   |
| PRIOR FILING DATE:   | 1996-07-12               |   |                   |
| PRIOR APPLICATION NUMBER:  | 60/001203                |   |                   |
| PRIOR FILING DATE:   | 1995-07-14               |   |                   |
| PRIOR APPLICATION NUMBER:  | 60/003,112               |   |                   |
| PRIOR FILING DATE:   | 1995-09-01               |   |                   |
| NUMBER OF SEQ ID NOS:  | 216                      |   |                   |
| SOFTWARE:  | PatentIn Ver. 2.0        |   |                   |
| SEQ ID NO  | 203                      |   |                   |
| LENGTH:  | 152                      |   |                   |
| TYPE:  | PRF                      |   |                   |
| ORGANISM:  | guinea pig               |   |                   |
| US-08-679-493A-203   |                          |   |                   |
| Query Match  |                          |   |                   |
| Best Local Similarity  | 5.7%;                    | Score 158;  | DB 4; Length 152; |
| Matches 44;  | Conservative 21;         | Mismatches 51;  | Indels 26; Gaps   |

```

Matches 43: Conservative 26: Mismatches 50: Indels 26: Gaps
OY 225 VHGCTIDFQVGYDLEYSYHLEGFNVSDDKDHLHDVQIVANDDLTSQCDNIGAKYPP-- 282
Db 14 VGGIHFQKANGVYVVKGRITGL-VEKGHGFHNEP-----CDNQGCTSAGRPFNPLS 67
OY 283 -----HEYHSELDDGLDHDHDDHGVNESHRSYWINIFGDDSVLGRSAIAHOR-DHLAK 336
Db 68 KKHGGRQDEEHVHVDLGNVYTAGADGVANVSIEDLSLSGANSIIGRTWVHEKFPDDLK 127
OY 337 -----SAKIACYIGRQQ 349
Db 128 GGNESYTGNAISRLAGVYIGAQ 152

RESULT 14
US-08-668-381A-5
; Sequence 5, Application US/08668381A
; Patent No. 5780024
; GENERAL INFORMATION:
; APPLICANT: Brown, Robert H.
; APPLICANT: Fishman, Paul S.
; APPLICANT: Francis, Jonathan W.
; APPLICANT: Hosler, Betsy A.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN
; TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668, 381A
; FILING DATE: 21-JUN-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/000,473
; FILING DATE: 23-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,164
; REFERENCE/DOCKET NUMBER: 00786/269001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 618 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-668-381A-5

Query Match 5.6%; Score 157.5; DB 1: Length 618;
Best Local Similarity 27.2%; Pred. No. 1.4e-06;
Matches 58: Conservative 36: Mismatches 64: Indels 55: Gaps 10.
OY 225 VHGCTIDFQVGYDLEYSYHLEGFNVSDDKDHLHDVQIVANDDLTSQCDNIGAKYPP- 282
Db 15 VGGIHFQKANGVYVVKGRITGL-----EGKHGFHNEP-----CDNQGCTSAGRPFNPL 68
OY 283 -----HEYHSELDDGLDHDHDDHGVNESHRSYWINIFGDDSVLGRSAIAHOR-DHLAK 335
Db 69 SRKHGPRQDEERHNVHVDLGNVYTAGKGVANVSIEDSVYLSGSDHCIIIGRTLVVHEKRADDL 128

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:33:20 ; Search time 21 seconds  
(without alignments)  
376,873 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 2794  
1 DGEQNDGQKKDDHDDHND.....MOGHEVESERVACVIGRA 497

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008

Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCOT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCOTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 168   | 6.0         | 150    | US-09-883-985-9     | Sequence 9, App1  |
| 2          | 166   | 5.9         | 151    | US-09-883-985-10    | Sequence 10, App1 |
| 3          | 165.5 | 5.9         | 152    | US-09-883-985-12    | Sequence 12, App1 |
| 4          | 160   | 5.7         | 1099   | US-09-881-654-4     | Sequence 4, App1  |
| 5          | 155.5 | 5.5         | 154    | US-09-904-987-6     | Sequence 6, App1  |
| 6          | 153.5 | 5.5         | 152    | US-09-883-985-7     | Sequence 3, App1  |
| 7          | 152.5 | 5.5         | 153    | US-09-883-985-3     | Sequence 7, App1  |
| 8          | 150   | 5.4         | 151    | US-09-883-985-8     | Sequence 8, App1  |
| 9          | 147   | 5.3         | 151    | US-09-883-985-5     | Sequence 11, App1 |
| 10         | 142.5 | 5.1         | 150    | US-09-883-985-11    | Sequence 2, App1  |
| 11         | 138   | 4.9         | 150    | US-09-883-985-2     | Sequence 4, App1  |
| 12         | 137   | 4.9         | 151    | US-09-883-985-6     | Sequence 9, App1  |
| 13         | 136.5 | 4.9         | 150    | US-09-883-985-4     | Sequence 21, App1 |
| 14         | 134   | 4.8         | 288    | US-09-964-899-21    | Sequence 7, App1  |
| 15         | 134   | 4.8         | 310    | US-09-925-297-795   | Sequence 1, App1  |
| 16         | 118   | 4.2         | 292    | US-09-864-761-37944 | Sequence 36, App  |
| 17         | 115   | 4.1         | 931    | US-09-864-761-37944 | Sequence 86, App1 |
| 18         | 112   | 4.0         | 725    | US-09-994-485-2     | Sequence 2, App1  |
| 19         | 110.5 | 4.0         | 409    | US-09-813-329-6     | Sequence 6, App1  |

|    |       |     |      |                     |                    |
|----|-------|-----|------|---------------------|--------------------|
| 20 | 109.5 | 3.9 | 544  | US-09-864-761-48016 | Sequence 48016, A  |
| 21 | 107.5 | 3.8 | 488  | US-09-840-787-17    | Sequence 17, App1  |
| 22 | 107.5 | 3.6 | 842  | US-09-798-831-8     | Sequence 8, App1   |
| 23 | 101.5 | 3.5 | 409  | US-09-813-329-2     | Sequence 2, App1   |
| 24 | 97    | 3.5 | 724  | US-09-994-485-4     | Sequence 4, App1   |
| 25 | 96    | 3.4 | 82   | US-09-864-761-33313 | Sequence 33313, A  |
| 26 | 95.5  | 3.4 | 2785 | US-09-801-574-8     | Sequence 8, App1   |
| 27 | 95    | 3.4 | 406  | US-09-813-329-4     | Sequence 4, App1   |
| 28 | 94    | 3.4 | 584  | US-09-815-242-5013  | Sequence 5013, AP  |
| 29 | 94    | 3.4 | 589  | US-09-815-242-10878 | Sequence 10878, A  |
| 30 | 91.5  | 3.3 | 1838 | US-09-879-248-8     | Sequence 8, App1   |
| 31 | 91    | 3.3 | 391  | US-09-764-870-328   | Sequence 328, App  |
| 32 | 90.5  | 3.2 | 835  | US-09-947-199-8     | Sequence 8, App1   |
| 33 | 90    | 3.2 | 477  | US-09-966-147-4     | Sequence 4, App1   |
| 34 | 90    | 3.2 | 822  | US-09-926-859A5     | Sequence 2, App1   |
| 35 | 89.5  | 3.2 | 835  | US-09-947-199-2     | Sequence 2, App1   |
| 36 | 88.5  | 3.2 | 216  | US-09-764-864-1451  | Sequence 1451, AP  |
| 37 | 88.5  | 3.2 | 0    | US-10-153-273-4     | Sequence 4, App1   |
| 38 | 88    | 3.1 | 222  | US-09-759-145-479   | Sequence 479, App  |
| 39 | 88    | 3.1 | 222  | US-09-780-665-479   | Sequence 479, App  |
| 40 | 88    | 3.1 | 222  | US-09-822-827-479   | Sequence 479, App  |
| 41 | 88    | 3.1 | 243  | US-09-864-761-43091 | Sequence 43091, A  |
| 42 | 87.5  | 3.1 | 156  | US-09-864-761-41679 | Sequence 141679, A |
| 43 | 87.5  | 3.1 | 677  | US-09-737-178-144   | Sequence 144, App  |
| 44 | 87.5  | 3.1 | 1143 | US-09-924-154-14    | Sequence 14, App1  |
| 45 | 87.5  | 3.1 | 1143 | US-09-924-154-14    | Sequence 14, App1  |

#### ALIGNMENTS

RESULT 1  
US-09-883-985-9  
; Sequence 9, Application US/09883985  
; Patent No. US20020081288A1  
; GENERAL INFORMATION:  
APPLICANT: YU, GUO-LIANG  
ROSEN, CRAIG A.  
FRASER, CLAIRE M.  
GOCAYNE, JEANNINE D.  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESS: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/883,985  
FILING DATE: 20-Jun-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/203,607  
FILING DATE: 02-DEC-1998  
APPLICATION NUMBER: US 08/722,050  
FILING DATE: 23-JAN-1997  
APPLICATION NUMBER: US 08/225,757  
FILING DATE: 11-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488, 1020003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 9:



Best Local Similarity 30.8%; Pred. No. 1.8e-07;  
Matches 44; Conservative 25; Mismatches 47; Indels 27; Gaps 6;

```

?
? CURRENT APPLICATION NUMBER: US/09/904,987
? CURRENT FILING DATE: 2001-07-12
? NUMBER OF SEQ ID NOS: 7
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO. 6
? LENGTH: 154
? TYPE: PRT
? ORGANISM: homo sapiens
? PUBLICATION INFORMATION:
? DATABASE ACCESSION NUMBER: NCBI ENTREZ /
? DATABASE ENTRY DATE: 2000-05-30
? RELEVANT RESIDUES: (1)..(154)
US-09-904-987-6

```

387  
Z / P00441  
E. DR 10. 100041b 1541.

```

1      RESULT 4
2      US-09-881-654-4
3      Sequence 4, Application US/098841654
4      Patent No. US20020146685A1
5      GENERAL INFORMATION:
6      APPLICANT: CHEN, David Y.
7      APPLICANT: ARCANDEL, Phillip
8      APPLICANT: TANDESKE, Laura
9      APPLICANT: GEORGE-MASCIMENTO, Carlos
10     APPLICANT: COLT, Doris
11     APPLICANT: MEDINA-SELBY, Angelica
12     TITLE OR INVENTION: IMMUNOSATS FOR ANTI-HCV ANTIBODIES
13     FILE REFERENCE: 2302-17039 / PPT7039.002
14     CURRENT FILING DATE: US/09/881,654
15     PRIOR FILING DATE: 2001-06-14
16     PRIOR APPLICATION NUMBER: 60/212,082
17     PRIOR FILING DATE: 2000-06-15
18     PRIOR APPLICATION NUMBER: 60/280,811
19     PRIOR FILING DATE: 2001-04-02
20     PRIOR APPLICATION NUMBER: 60/280,867
21     PRIOR FILING DATE: 2001-04-02
22     NUMBER OF SEQ ID NOS: 7
23     SOFTWARE: PatentIn Ver. 2.0
24     SEQ ID NO 4
25     LENGTH: 1099
26     TYPE: PRT
27     ORGANISM: Artificial Sequence
28     FEATURE:
29     OTHER INFORMATION: Description of Artificial Sequence:
30     US-09-881-654-4
31     MEPA 7.1.1

```

RESULT 6  
US-09-883-985-7  
Sequence 7, Application US/09883985  
Patent No. US20020081288A1  
GENERAL INFORMATION:  
APPLICANT: YU, GUO-LIANG  
ROSEN, CRAIG A.  
FRASER, CLAIRE M.  
GOCAYNE, JEANNINE D.  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P L L C  
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
CITY: WASHINGTON

```

Query Match      5.7%; Score 160; DB 10; Length 1099;
Best Local Similarity 26.4%; Pred. No. 7,78-06;
Matches 46; Conservative 35; Mismatches 55; Indels 38; Gaps 7;

OY   225 VHGTHDFKQ-VGYGDLDEVSHLEGFNVSDDHDKHDHVDQIANDLTSGCDNIGAKYPD-282
      |.|.:|.|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|
Db    15 VQGIIINEQESNGPKVMWSIKGLT-----EGLAGHHVEHFSGDNTAGCTSSAPPHENPL 68
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

OY   283 -----HEDYSELGDLDGIHDDDHGVNESHRYSWNIIFGDSVLGRSTAIHOR-DLHL 335
      |.|.:|.|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|
Db    69 SRKHGPRDERRHVGDIIGNVTADKCGADVSTIEDSVISLSDBHCITIGTLVHERKADLG 128
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

OY   336 K-----SARIACCVIGRGOSHPEIVYHRAKCVCVARNTSEGTGLHHVSG 377
      |.|.:|.|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|.|.:|
Db    129 KGCEESTKTGNAGSRILAGVIGIAQN---LNMSGCNCSTIYG-----HITS 171
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
US-09-904-987-6
Sequence 6, Application US/09904987
Patent No. US20020037908A1
GENERAL INFORMATION:
APPLICANT: NO. US20020037908A1acetyl, INC.
TITLE OR INVENTION: Methods and Compositions for Controlling Pathological and Prepat
TITLE OF INVENTION: Protein Assembly or Aggregation
FILE REFERENCE: 42108/26146
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.3  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/883,985  
FILING DATE: 20-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/203,607  
FILING DATE: 02-DEC-1998  
APPLICATION NUMBER: US 08/722,050  
FILING DATE: 23-JAN-1997  
APPLICATION NUMBER: US 08/225,757  
FILING DATE: 11-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: STEEPE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.1020003  
TELECOMMUNICATIONS INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 152 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-883-985-7

Query Match 5.5%; Score 153.5; DB 10; Length 152;  
Best Local Similarity 28.7%; Pred. No. 2e-06;  
Matches 41; Conservative 30; Mismatches 45; Indels 27; Gaps 5;

OY 225 VHGTDKQ-VGYGDLVSYHLEGFNVSDDKHLADVOIYANDLTSGCDNLGAKYDP-282  
DB 13 VQGIINEQKESGPKVYVMSIKGLT-----EGLHGFVHFEFGDNTAGCTSAGPHENPL-66  
OY 283 -----HEDYHSELGLGDIHDDHGVNESHVSWINIFGDDSVLGRSLAHOR-DHLH-335  
DB 67 SRKHGPKDERHVGDLGNYTADKDGADVSTEDSVISLSGDHCIIGRITLVHEKADDLG-126  
OY 336 K-----SAKIACCVIG-346  
DB 127 KGNNESTKTGNMAGSLACGVIG-149

RESULT 7  
US-09-883-985-3  
Sequence 3, Application US/09883985  
Patent No. US20020081288A1  
GENERAL INFORMATION:  
APPLICANT: YU, GUO-LIANG

ROSEN, CRAIG A.  
FRASER, CLAIRE M.  
GOCAYNE, JEANNINE D.  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P. L. L. C.  
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/883,985  
FILING DATE: 20-Jun-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/203,607  
FILING DATE: 02-DEC-1998  
APPLICATION NUMBER: US 08/722,050  
FILING DATE: 23-JAN-1997  
APPLICATION NUMBER: US 08/225,757  
FILING DATE: 11-APR-1994

ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.1020003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2540  
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 153 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-883-985-3

Query Match 5.5%; Score 152.5; DB 10; Length 153;  
Best Local Similarity 29.4%; Pred. No. 2.4e-06;  
Matches 46; Conservative 28; Mismatches 52; Indels 35; Gaps 9;

OY 357 RACVVRPNTSEGLHHVYSGITP-EOTPGGSHMTADLKGFNVEDLSHHKGVQLH-415  
DB 2 KACVVM--TGTAG--VKGVKFTQETDNGPVVHNEFSG-----LAKGKHGFVHE-48  
OY 416 WGDMSHGCHSLGRMYHGHDAHP---KRPDGLDVI--DSDHGIHVSTRTFDHLN-V-467  
DB 49 FGDTTNGCTAGAHFNPTKQEHGAPEDSIRHYGDIGNVVAGADGNNAVYMTDKLISLNGS-108  
OY 468 EDLNARSLVIMQ-----GHEVE-----SERVACCVIGRA-497  
DB 109 HSLGRSMVTHENEDDLGRGHLSKVGTGNAGRLACGVGLA-151

RESULT 8  
US-09-883-985-8  
Sequence 8, Application US/09883985  
Patent No. US20020081288A1  
GENERAL INFORMATION:  
APPLICANT: YU, GUO-LIANG

ROSEN, CRAIG A.  
FRASER, CLAIRE M.  
GOCAYNE, JEANNINE D.  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P. L. L. C.  
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
STATE: WASHINGTON  
COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/883,985  
FILING DATE: 20-Jun-2001  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/203,607  
FILING DATE: 02-DEC-1998  
APPLICATION NUMBER: US 08/722,050  
FILING DATE: 23-JAN-1997  
APPLICATION NUMBER: US 08/225,757  
FILING DATE: 11-APR-1994

ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.  
REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.1020003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2540  
TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 151 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-883-985-8

Query Match 5.4%; Score 150; DB 10; Length 151;

```

Best Local Similarity 30.3%; Pred. No. 3,9e-06; Indels 26; Gaps 4;
Matches 43; Conservative 21; Mismatches 52;

OY      225 VHGIDPKQVGYDDLEVSYLEGFWNSDDHKDHLHDVOIYNAGSDLTSGCDNLGAKYDP-- 282
| | | | | | | | : : : | | | | | | | | | | | | | | | | | | | | | |
Db      13 VSGVIETQVGAPPTVYNGNISGL-----KFGHGFHVHALGDNTGCGSTGPHTNPG 66
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      283 -----HEDYISELDIGDIHHDDHGVNVESHRYSHINIFGDSVLGSIAIH-QRPHLAK 336
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      67 KKHGAFDEYVRHAGDTIGNITGEVDGSAFTYTDKOIPLETGHOSITIGRAVVHADPDLLK 126
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      337 -----SAKIACCVIG 346
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      127 GGHELSKSTGMAGRIACGIIG 148
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-883-985-5
Sequence 5, Application US/09883985
Patent No. US20020081288A1
GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
ROSEN, CRAIG A.
FRASER, CLAIRE M.
GOCCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEEFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-883-985-5

Query Match 5.3%; Score 147; DB 10; Length 151;
Best Local Similarity 30.3%; Pred. No. 7.1e-06;
Matches 43; Conservative 21; Mismatches 52; Indels 26; Gaps 4;
225 VHGIDPKQVGYDDLEVSYLEGFWNSDDHKDHLHDVOIYNAGSDLTSGCDNLGAKYDPH- 283
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      13 VKGIPTFTHGNCATVITGVSGL-----RPLGLGHFHVALDDNTNGCMSTGPHEHNPDG   66
Oy      284 -----EDYSELGDGDIDHDDHCVVNESHRYSWINIFGDDSVLGSRJAIH-QRHLLXK   336
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      67 KTHGAEPDANRRHAGDGGNIIVGDDGTAFITTDISOIPLSPGNISVGRAIVAHADPDDLK    126
Oy      337 -----SAKIACCYIG 346
           ::||::|||
Db      127 GGHELSTGNCAGRVACGII 148

RESULT 10
; US-09-883-985-11
; Sequence 11, Application US/09883985
; Patent No. US20020081288A1
; GENERAL INFORMATION:
APPLICANT: YU, GUO-LIANG
            ROSEN, CRAIG A.
            FRASER, CLAIRE M.
            GOCAYNE, JEANNINE D.
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/883,985
FILING DATE: 20-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/203,607
FILING DATE: 02-DEC-1998
APPLICATION NUMBER: US 08/722,050
FILING DATE: 23-JAN-1997
APPLICATION NUMBER: US 08/225,757
FILING DATE: 11-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.1020003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 150 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-883-985-11

Query Match          5.1%; Score 142.5; DB 10; Length 150.
Best Local Similarity 28.2%; Pred. No. 1.7e-05;
Matches 40; Conservative 25; Mismatches 50; Indels 27; Gaps 5;

Oy      225 VHGTFDRKYQGYGDLLEVSYHLGCFNVSDDKHKHLHDVOIVANGDTLTSCDNILGAKYPDPE 284
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      13 VKGVAREEQQDDDDVEKGIEELT-----DGNNGFIHVFGDNTGCCLSLAPPHFPQN   66
Oy      285 DYHS-----ELGDLGDIHHDDHGVNESHRYSWINIFGDDSVLGSRJAINHQR-DHLRK 336
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      67 KNHSPKDADRHHGIDGAV-TAEGVAGDFNFDPDISLKGRSRIIGRTAVVHKODDLK    125

```

OY 337 -----SAKIACCVIG 346  
Db 126 GGDESLKTNAGCRLACGVIG 147

RESULT 11  
US-09-883-985-2  
Sequence 2, Application US/09883985  
Patent No. US20020081288A1

GENERAL INFORMATION:  
APPLICANT: YU, GUO-LIANG

ROSEN, CRAIG A.  
FRASER, CLAIRE M.

GOCAYNE, JEANNINE D.  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4

NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
CITY: WASHINGTON

STATE: DC  
COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/883,985  
FILING DATE: 20-Jun-2001

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/203,607  
FILING DATE: 02-DEC-1998

APPLICATION NUMBER: US 08/722,050  
FILING DATE: 23-JAN-1997

APPLICATION NUMBER: US 08/225,757  
FILING DATE: 11-APR-1994

ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.1020003

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2540

TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 255 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-883-985-2

Query Match 4.9%; Score 138; DB 10; Length 255;  
Best Local Similarity 33.3%; Pred. No. 8.7e-05;

Matches 36; Conservative 19; Mismatches 33; Indels 20; Gaps 4;

OY 409 HGVOLEHMGDSHGCHSLGRMY-----HGHDADHPKRPGLDGLVDYIDDSHGIVHSRTF 462  
Db 108 HGLVHVGCGDITNCGNHNFPDASHGPGQSDSDRHR-GDLGNVRADADGRIAFMED 166

OY 463 DHLVNEEDLNARSLVIMO-----GGHEV-----ESERVACCVIGRA 497  
Db 167 EOLKMWVYIARSLIIDGEGDGLGGHPLSKITNGSGERLACGIARS 214

RESULT 12  
US-09-883-985-4  
Sequence 4, Application US/09883985

Patent No. US20020081288A1  
GENERAL INFORMATION:  
APPLICANT: YU, GUO-LIANG

ROSEN, CRAIG A.  
FRASER, CLAIRE M.

GOCAYNE, JEANNINE D.  
TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4

NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600  
CITY: WASHINGTON

STATE: DC  
COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/883,985  
FILING DATE: 20-Jun-2001

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/203,607  
FILING DATE: 02-DEC-1998

APPLICATION NUMBER: US 08/722,050  
FILING DATE: 23-JAN-1997

APPLICATION NUMBER: US 08/225,757  
FILING DATE: 11-APR-1994

ATTORNEY/AGENT INFORMATION:  
NAME: STEFFE, ERIC K.

REGISTRATION NUMBER: 36,688  
REFERENCE/DOCKET NUMBER: 1488.1020003

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 150 amino acids

TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-883-985-4

Query Match 4.9%; Score 137; DB 10; Length 150;  
Best Local Similarity 29.5%; Pred. No. 5.2e-05;

Matches 44; Conservative 16; Mismatches 55; Indels 34; Gaps 6;

OY 54 VHGSIELSKGHGAVYLELHNGFNTSEDDHNNGLHLMGDSACDSIGELYNANP 113  
Db 13 VCGTTHFEKAGDTYV-----VTGSTITGLTEGD--HGFMHOGSDWTGCTGAGHFNPLS 65

OY 114 EKRRAD-----GDLGDLVDDNGVYNEVNHAYMLDIG--TAPNTEALIGSMITLQ 163  
Db 66 KKHGKDEERHVGDLGVATADKNGV-----AYDYIDPLISLSGEYSITIGRTMYVNE 118

OY 164 -----GSHTDADTPASRIACCVIG 182  
Db 119 KPDDLGRGNESESTYTGNAAGSRLACGVIG 147

RESULT 13  
US-09-883-985-6  
Sequence 6, Application US/09883985

Patent No. US20020081288A1  
GENERAL INFORMATION:

APPLICANT: YU, GUO-LIANG  
ROSEN, CRAIG A.

FRASER, CLAIRE M.

Mon Dec 2 12:20:03 2002

us-09-868-760-7.rapb

Page 7

```

      GOCAYNE, JEANNINE D.
      TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
      NUMBER OF SEQUENCES: 16
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: STERNER, KESSLER, GOLDSTEIN & FOX P.L.L.C.
      STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
      CITY: WASHINGTON
      STATE: DC
      COUNTRY: USA
      ZIP: 20005
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      OPERATING SYSTEM: IBM PC compatible
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/883,985
      FILING DATE: 20-Jun-2001
      CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 09/203,607
      FILING DATE: 02-DEC-1998
      APPLICATION NUMBER: US 08/722,050
      FILING DATE: 23-JAN-1997
      APPLICATION NUMBER: US 08/225,757
      FILING DATE: 11-APR-1994
      ATTORNEY/AGENT INFORMATION:
      NAME: STEFFE, ERIC K.
      REGISTRATION NUMBER: 36,688
      REFERENCE/DOCKET NUMBER: 1488.1020003
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (202) 371-2600
      TELEFAX: (202) 371-2540
      INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 151 amino acids
      TYPE: amino acid
      STRAND: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-883-985-6

Query Match
Best Local Similarity 4.98; Score 136.5; DB 10; Length 151;
Matches 44; Conservative 21; Mismatches 49; Indels 27; Gaps 6;
QY 227 GTIDKQVGV-DLEVYHLEGNNVSDDKHLDHVDQIYANGDLTSGCDNLGAKYDPHED 285
DB 13 GTVEEEOSSGTPKVSQVCGL-----AKLHGFEVHEFGDVTNCGSSGPHFNPYK 66
QY 286 YHS-----ELDIDGDIHDDHGVVNSHRYSWINIFGDDSVLGRSIAIH-----QR 331
DB 67 EHGAPVDENHLDGDIENATGDCPTKYNTIDSKITLFGADSIIGRTVVVHADDDLGQ 126
QY 332 DR-LHRS-----AKIACVYG 346
DB 127 GHELSKSTGNAGARICGVIG 147

RESULT 14
US-09-964-899-21
Sequence 21, Application US/09964899
Patent No. US2002017446A1
GENERAL INFORMATION:
APPLICANT: Cohen, Dalia et al.
TITLE OF INVENTION: Identification of Genes Involved in
      CURENCE: 4-31612 A
      CURRENT APPLICATION NUMBER: US/09/964,899
      PRIOR FILING DATE: 2001-09-27
      PRIOR APPLICATION NUMBER: 60/236,893
      PRIOR FILING DATE: 2000-09-29
```

```

      PRIOR APPLICATION NUMBER: 60/236,309
      PRIOR FILING DATE: 2001-06-14
      NUMBER OF SEQ ID NOS: 53
      SOFTWARE: FastSeq for Windows Version 4.0
      SEQ ID NO: 21
      LENGTH: 288
      TYPE: PRT
      ORGANISM: Homo Sapien
US-09-964-899-21

Query Match
Best Local Similarity 4.88; Score 134; DB 9; Length 288;
Matches 35; Conservative 19; Mismatches 34; Indels 20; Gaps 4;
QY 409 HGVQLHEWGMDSHCHSLGRMY-----HGHDADHPKRPGLGVDYIDSHGIVSTRFE 462
DB 141 HGLVHQYGDLTNNCSGNNHFPDGAHGGPOSDRHR-GDLGNVRADADGRAIFRMD 199
QY 463 DHLNVEDLNARSLVIMQ-----GGHEV-----ESERVACVIGRA 497
DB 200 EQLKWDVIGRSLIIDEGEDDLGRGHPLSKITNGSGERLACGIARS 247

RESULT 15
US-09-925-297-795
Sequence 795, Application US/09925297
Patent No. US20020081659A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
      CURRENT APPLICATION NUMBER: US/09/925,297
      PRIOR FILING DATE: 2001-08-10
      PRIOR APPLICATION NUMBER: PCT/US00/05989
      PRIOR FILING DATE: 2000-03-08
      PRIOR APPLICATION NUMBER: 60/124,270
      NUMBER OF SEQ ID NOS: 928
      SOFTWARE: Patentin Ver. 2.0
      SEQ ID NO: 795
      LENGTH: 310
      TYPE: PRT
      ORGANISM: Homo sapiens
      NAME/KEY: SITE
      LOCATION: (772)
      OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-795

Query Match
Best Local Similarity 4.88; Score 134; DB 10; Length 310;
Matches 35; Conservative 19; Mismatches 34; Indels 20; Gaps 4;
QY 409 HGVQLHEWGMDSHCHSLGRMY-----HGHDADHPKRPGLGVDYIDSHGIVSTRFE 462
DB 163 HGLVHQYGDLTNNCSGNNHFPDGAHGGPOSDRHR-GDLGNVRADADGRAIFRMD 221
QY 463 DHLNVEDLNARSLVIMQ-----GGHEV-----ESERVACVIGRA 497
DB 222 EQLKWDVIGRSLIIDEGEDDLGRGHPLSKITNGSGERLACGIARS 269

Search completed: November 29, 2002, 12:37:34
Job time : 24 secs
```

THIS PAGE BLANK (USPTO)





[illegible]

RESULT 2  
T45059  
Hypothetical protein Y39B6b.gg [imported] - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T45059  
R:Wilson, R.; Altschough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton, L.; Cullen, A.; Fulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Jones, B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopka, A.; Saunders, D.  
Nature 368, 32-38, 1994  
A:Authors: Showkneen, R.; Sims, M.; Smaildon, N.; Smith, A.; Smith, M.; Sonhammer, E.; Slock, L.; Wilkinson-Spoat, J.; Wohldman, P.  
A:Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.  
A:Reference number: S45331; MUID:94150718; PMID:7906398  
A:Accession: T45059  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-735 <WILD>  
A:Cross-references: EMBL:AL132896; NID:96434440; PIDN:CAB60938.1; PID:96434473  
A:Experimental source: clone Y39B6b  
C:Genetics:  
A:Map position: 3  
A:Introns: 18/1; 69/1  
A:Note: Y39B6b.gg

[illegible]

|    |     |         |     |
|----|-----|---------|-----|
| QY | 464 | HLNVED  | 469 |
|    |     | ! : ! : |     |
| Db | 706 | HDDKEN  | 711 |

### RESULT 3

superoxide dismutase (EC 1.15.1.1) (Cu-zn) - soybean  
 N:Alternate names: Cu/Zn-SOD  
 C:Species: Glycine max (soybean)  
 C:Date: 18-Jun-1998 #sequence-revision 10-Jul-1998 #text-change 20-Apr-2000  
 C:Accession: JMW0084  
 R:Arachita, M.; Nong, V.H.; Kadokura, K.; Kimura, K.; Ueda, K.; Fukazawa, C.  
 Biosci. Biotechnol. Biochem. 62, 1018-1021, 1998  
 A:Title: Molecular cloning and expression patterns of Cu/Zn-superoxide dismutases in  
 A:Reference number: JMW0084; MUID:98312043; PMID:9648237  
 A:Accession: JMW0084  
 A:Molecule type: mRNA  
 A:Residues: 1-152 <RNA>  
 A:Function:  
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxy  
 A:Superfamily: superoxide dismutase (Cu-zn)  
 C:Keywords: copper; homodimer; metalloprotein; oxidoreductase; zinc  
 E:45-47/62-119/Binding site: copper (His) #status predicted  
 E:56-145/Disulfide bonds: #status predicted  
 E:62-70/79-82/Binding site: zinc (His, His, Asp) #status predicted  
 E:62/Active site: Arg #status predicted

|                       |                 |                   |           |            |
|-----------------------|-----------------|-------------------|-----------|------------|
| Query Match           | 6.4%            | Score 178         | DB 2      | Length 152 |
| Best Local Similarity | 33.1%           | Pred. No. 4,8e-06 |           |            |
| Matches 47            | Conservative 23 | Mismatches 46     | Indels 26 | Gaps 4     |

  

|    |     |  |     |
|----|-----|--|-----|
| QY | 225 | VHGITDRCQVGYGDDEVSHTLDEGFVNSDDKHCHLDVQIVANGSLTGGCNGTGAAXYDPP | 283 |
|    |     |  |     |
| Db | 14  | VGTGTFITGCGNPTVYTGSLAGL-----KFLGLGFVHNLGDTTNGCLSTGHNRPNN     | 67  |
|    |     |  |     |
| QY | 284 | -----EDYSELGADLDIHDHDDHGVYNESHRYSWINIFGDSVLSGSLAH-QRDHLK     | 336 |
|    |     |  |     |
| Db | 68  | NEHGAPEDENRHHAGDIGNVNGDDGVTSITDSQIPLTGPNSTITGAAVYVHADSDDLK   | 127 |
|    |     |  |     |
| QY | 337 | -----SAKITACCVIG   | 346 |
|    |     | ::: ::   |     |
| Db | 128 | GGHLSKTTGNAGGRVACGIIG  | 149 |

```

RESULT 4
A14373 histidine-rich calcium-binding protein precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 05-Nov-1999
C:Accession: A14373
R:Hotmann, S.L.; Goldstein, J.L.; Orth, K.; Moosaw, C.R.; Slaughter, C.A.; Brown, M.S.
J. Biol. Chem. 264, 18083-18090, 1989
A:Title: Molecular cloning of a histidine-rich Ca(2+)-binding protein of sarcoplasmic
A:Reference number: A14373; MUID:90036684; PMID:2808365
A:Accession: A14373
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1 852 <NO>
A:Cross-References: GB:J05080; NID:9165099; PIDN:AAA1279.1; RID:9165100
C:Keywords: calcium binding

Query Match          6.4%; Score 178; DB 2; Length 852;
Best Local Similarity 22.2%; Pred. No. 4.3e-05;
Matches 126; Conservative 54; Mismatches 176; Indels 212; Gaps 38;

Q1 1 DGEQCNQGKNDHNDH-----HDHNDHDDDDDETMTYAAQCEKRPYPMASLLNNHNG 56
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 237 DDDDEDDSTESDHOAHNRHGRGHEEEDDDDE-----GDSTESDRHOA---NRH--- 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Q2 57 SIELSGKH-----GAVYLLELVL-----GNTSEDHD-----HNHGL 90
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
Db 286 -----RGHREDEDDDDGGDSTESDRHQANRHGRHEEDDDDDDDGGDSTESDRHQ 339
Oy 91 HLM-----LGMASAGDSTIGELYNH-----PEKNADPRGLDYLNDGGVYNEH 137
Db 340 HNRHGRHEDEDDDDGGDSTESDRHQANRHGRHEDEDDDDDDGGDSTESDR-----HQAN 396
Oy 138 HYH-----LIDGTAPRTALIGHSTIL-----QGSHTDADT-PASR1 176
Db 397 RHNGHREDEDDDDGGDSTESDRHQANRHGRHEDEDDDDGGDSTESDRHQANRH 454
Oy 177 ACCVIGHKARPE--TAAALHHELE-----EDKTE-HYAH----- 208
Db 455 R-----GHGEEDDDDDGGDSTESDRHQANRHGRHEDEDDDDGGDSTESDRHQANRH 510
Oy 209 ---CDVRSNTHOPKALHNHV--HGTIDFKOVGYDLEVSYLEGFNVSDDKHDLNDVOI 263
Db 511 KEAEVTSDEH-----HHNVPRHG-----HOGHGDKEGE--EGVST-----DHMHQVPR 553
Oy 264 YAN-----GDLTSGCDNLGAKYPRHEDYHSELGDLG-DIHDDHGVV--NESHRYSMINI 315
Db 554 HAHNGGEGEEGDEELTYKAGHNVAHPRPGHRSREGHAEEHOTEVGRHNGH----- 607
Oy 316 FGDDVIGRSIAIHORDHLHKSAKIAACVIGRGOSHPETIYRAKCVVAPRTSTGLHNHV 375
Db 608 -----MGDTDSAEGRHASSP-----RQGNP-----PBDT-VNHN- 638
Oy 376 SGSTFE-----QTPG-----GSTHMTADLKGFNVSDDLSHNRGCVOLHEMGDS 420
Db 639 RGSLEKEVGPSPGPAKYDGSRYKRGSGSEEBEOKG-----TTH-HSLEDEDEBERG 690
Oy 421 HGCHSLGRMYHGHDADHPKRPGLADY 448
Db 691 HG-RSLSQ-----EDQEEEDRGESAKY 712

RESULT 5
552859
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jun-1995 #sequence_revision 21-Jul-1995 #text_change 05-May-2000
C:Accession: I48335; S52859
R:Carlsson, L.M.; Jonsson, J.; Edlund, T.; Marklund, S.L.
Proc. Natl. Acad. Sci. U.S.A. 92, 6264-6268, 1995
A:Title: Mice lacking extracellular superoxide dismutase are more sensitive to hyperoxia
A:Reference number: I48335; MUID:95327627; PMID:7603981
A:Accession: I48335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-251 <RES>
A:Cross-references: EMBL:X84940; NID:9695628; PIDN:CAA5935.1; PID:9695629
C:Genetics:
A:Gene: SOD3
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:148,130,145/Binding site: copper (His) #status predicted
F:145,153,156,159/Binding site: zinc (His, His, His, Asp) #status predicted
F:218/Active site: Arg #status predicted

Query Match 6.2%; Score 174.5; DB 2; Length 251;
Best Local Similarity 31.1%; Pred. No. 1.e-05;
Matches 56; Conservative 22; Mismatches 57; Indels 45; Gaps 11;

Oy 199 EEDKTEHYACDVRSNTHOPKAL-----HHNVGTIDFKOVGYD-LFVSYLEGFNVSDD 253
Db 67 EVDAAEHNAICRY-----QPSATLPRDPOQTGLVLFQQLDPSRLFAVYSLGFPRAEON 121
Oy 254 HKD-----HLNDVOIYANGDITSGCDNLGAKYPRHEDYHSELGDLADIDDDHG--VYNES- 307
Db 122 ASNRALIHVEF-----GDLSSQCDSTGPHYMPMEVPRHQ-----HPSGFGFVVRNQ 169
Oy 308 ---HRS-WLNIFGDDSVIGRSIAIHQ-RDLHLKSA-----KIACCVIGRGOS 350
```

```
Db 170 LMRHRYGLASLAGPNSHISIGRSVYVNHAGEDDLGKGNGASLQNGMRRLACCVYTSSS 229

RESULT 6
106570
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: T06570
R:White, D.A.; Zilinskas, B.A.
Plant Physiol. 96, 1391-1392, 1991
A:Title: Nucleotide sequence of a complementary DNA encoding pea cytosolic copper/zinc
A:Reference number: 215766
A:Accession: T06570
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-152 <WHI>
A:Cross-references: EMBL:W63003; NID:9169069; PIDN:AAA3659.1; PID:9169070
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:45,47,62,119/Binding site: copper (His) #status predicted
F:62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted
F:112/Active site: Arg #status predicted

Query Match 6.2%; Score 173.5; DB 2; Length 152;
Best Local Similarity 31.4%; Pred. No. 1.e-05;
Matches 49; Conservative 26; Mismatches 48; Indels 33; Gaps 5;

Oy 211 VRSNTHOPKALHNHVHGTIDFKOVGYDLEVSYLEGFNVSDDKHDLNDVOIYANGDYL 270
Db 7 VLSNSNE-----VSGTIFSEQNGNGPTVTYGTLAGL-----KPLHGFHIALGDTT 53
Oy 271 SGCDDNLGAKYPR-----EDYHSELGDLIDHDDHGVYNSHRSWVNIINGDSDVIG 323
Db 54 NCGISITGPHNPNGKEHGAPEDETHAGDLGNINVGDDGVFTITNHLPLGTNSIIG 113
Oy 324 RSLAIH-QBDHLHK-----SAKIKCVIG 346
Db 114 RAVVYHADPDDLGKSGHELSKTTGNAGRVACGIIIG 149

RESULT 7
A49097
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) B precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 05-May-2000
C:Accession: A49097; S34053; S37408
R:Williams, J.; Zwijsen, A.; Siegers, H.; Nicolate, S.; Bettadapura, J.; Raymackers, J.
J. Biol. Chem. 268, 24614-24621, 1993
A:Title: Purification and sequence of rat extracellular superoxide dismutase B secret
A:Reference number: A49097; MUID:94043314; PMID:8227019
A:Accession: A49097
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-244 <WHI>
A:Cross-references: EMBL:Z24721; NID:9404250; PIDN:CAA80849.1; PID:9404251
R:Perry, A.C.F.; Jones, R.; Hall, L.
Biochem. J. 293, 21-25, 1993
A:Title: Isolation and characterization of a rat cDNA clone encoding a secreted super
A:Reference number: S34053; MUID:93319511; PMID:8328962
A:Accession: S34055
A:Molecule type: mRNA
A:Residues: 1-234, 'W', 236, 'W', 238-244 <PER>
A:Cross-references: EMBL:X68041; NID:9394738; PIDN:CAA48177.1; PID:9394739
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:121,123,138,168/Binding site: copper (His) #status predicted
```



A:Accession: A92281  
A:Molecule type: protein  
R:Residues: 2-55,'D','57-92','D','94-154 <STP>  
R:Remazani Rad, M.; Kirchraeth, L.; Hollenberg, C.P.  
Submitted to the Protein Sequence Database, September 1995

A:Reference number: S57111  
A:Accession: S57125  
A:Molecule type: DNA  
A:Residues: 1-154 <RAM>

R:Horiuchi, S  
A:Cross-references: EMBL:D49604; NID:g1015811; PIDN:CAAB9634.1; PIDs:g1015812; GSPDB:GNOD  
submitted to the Protein Sequence Database, March 1992

A:Reference number: A40093  
A:Accession: A40093  
A:Molecule type: protein  
A:Residues: 2-7,'KL','10-42','V','44-55','D','57-88','L','90-91','AD','94-100','Q','102','K','104-107'  
C:Genetics:  
A:Gene: SGD:SODI; MIPS:YJR104c  
A:Cross-references: SGD:S003865; MIPS:YJR104c  
A:Map position: 10R  
C:function:

C:description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
C:superfamily: superoxide dismutase (Cu-Zn)  
C:keywords: copper; metalloprotein; oxidoreductase; zinc  
F:2-154/Product: superoxide dismutase (Cu-Zn) #status experimental <MAT>  
F:47,49,64,121/Binding site: copper (His) #status predicted  
F:58-147/Dissulfide bonds: #status experimental  
F:64,72,81,84/Binding site: zinc (His, His, Asp) #status predicted  
F:144/Active site: Arg #status predicted

Query Match            6.0%   Score 167: DB 1: Length 154;  
Best Local Similarity   30.1%; Pred. No. 3e-05;  
Matches      43; Conservative     26; Mismatches    48; Indels       26; Gaps          5;

Oy      225 VHGIDPKOVCYGD -LEVSYLHEGFNSDDKHNDHVDQIRANDPLSGGDNIGAKAYDPH 283  
| | : || : |||| : |||| : |||| : |||| : |||| : |||| : ||||  
Db      14 VSGLVKKEQAASEEPITTVETELAGSNPAERGFHHNE----GDATNGCVSAGPRHFNF 68

Oy      284 EDYS-----ELGGLDIHDDHCYVNESHRYRWINIFEDSDVLGRSLAIHQ-RDHHL 335  
| | : || : |||| : |||| : |||| : |||| : |||| : ||||  
Db      69 KTHIGAFTDEVRIYGVDMGYKTDEMGVAKGSRDSLILKITPSIVGSVVIIHAGDDDLG 128

Oy      336 K-----SAKIACCYTG 346  
| | : || : |||| : |||| : |||| : |||| : |||| : ||||  
Db      129 KGDTESLKTTGNAGRRCVGIG 151

RESULT 11  
DSHOXZ  
Superoxide dismutase (EC 1.15.1.1) (Cu-Zn) 1 [validated] - horse  
C:Species: Equus caballus (domestic horse)  
C>Date: 02-Apr-1982 #sequence revision 09-Aug-1997 #text\_change 20-Apr-2000  
C:Accession: JCS5215; A00515  
R:de la Rúa-Domech, R.; Wiedmann, M.; Mohamed, H.O.; Cummins, J.F.; Divers, T.J.; Be-

Ene 178, 83-88, 1996  
A>Title: Equine motor neuron disease is not linked to Cu/Zn superoxide dismutase mutation  
A:Reference number: JCS5215; MUId:97080551; PMID:8921896  
A:Accession: JCS5215  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-154 <DEA>

A:CROSS-references: GB:U38956; NUD:g1228115; PIDN:AAC48662.1; PID:g1228116  
B:Experimental source: leukocyte  
R:Bierich, K.; Ammer D  
J Biol Chem 256, 11545-11551, 1981

A>Title: Amino acid sequence of copper-zinc superoxide dismutase from horse liver.

A:Reference number: A00515; MUId:82052979; PMID:728616  
A:Accession: A00515  
A:Molecule type: protein  
A:Residues: 2-154 <LER>  
C:Genetics:  
A:Gene: sod1  
C:function:

```

A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: acetylated amino end; copper; metalloprotein; oxidoreductase; zinc
F:2/Modified site: acetylated amino end (Ala) (in mature form) #status experimental
F:47/49,64,121/Binding site: copper (His) #status predicted
F:58-117/Disulfide bonds: #status predicted
F:64,72,81,84/Binding site: zinc (His, His, Asp) #status predicted
F:114/Active site: Arg #status predicted

Query Match          5.9%; Score 164.5; DB 1; Length 154;
Best Local Similarity 31.0%; Pred. No. 4,6e-05;
Matches 45; Conservative 26; Mismatches 43; Indels 31; Gaps 6;

QY      225 VHGTFDKGVGYGDLEVSYLEGF--NVSDHKDHLHDVOIYANGDLTSGCDNLGAKYD 281
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       15 VHGIHFEEQQOEGRPV---LKGFTGLTKDGDFHVFHFEF-----GNPTGGCTTAAMFN 66

QY      282 P-----HEDYSESLGDIIDDDHGVAVESRYSINIFGDSYLGRSIAIHQR-DH 333
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       67 PLSKKHGGPDEERHHVGDTIGNTADENCKADVDKMDSYISLGSKHSTIGRTVMVHEKKOD 126

QY      334 LHK-----SAKIALCVIG 346
        |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       127 LGKGGNEESTKTGMAGSRILACGVIG 151

RESULT 12
S07007
Superoxide dismutase (EC 1.15.1.1) (Cu-Zn) 4, cytosolic [validated] - maize
C:Species: Zea mays (maize)
C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000
C:Accession: S07007; S20569; S72234
R:Canham, R.E.; Scandalios, J.G.
Mol. Gen. Genet. 219, 1-8, 1989
A>Title: Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in
A:Reference number: S07007; MUID:90136495; PMID:2482436
A:Accession: S07007
A:Molecule type: mRNA
A:Residues: 1-152 <CAN>
A:Cross-references: EMBL:X17565; NID:922483; PIDN:CA857992.1; PID:96018682
A:Experimental source: Inbred line W64A
A:Accession: S20569
A:Molecule type: protein
A:Residues: 2-20 <CAN2>
R:Kernodle, S.P.; Scandalios, J.G.
Genetics 144, 317-328, 1996
A>Title: A comparison of the structure and function of the highly homologous maize an
A:Reference number: S72234; MUID:97032947; PMID:8678655
A:Accession: S72234
A:Molecule type: DNA
A:Residues: 1-124 <KEP>
A:Cross-references: EMBL:U34726; NID:g1574937; PIDN:AAB49912.1; PID:g1574938
A:Gene: sod4
A:Introns: 26/1; 60/1; 92/1; 102/3
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; cytosol; metalloprotein; oxidoreductase; zinc
F:2-152/Product: superoxide dismutase (Cu-Zn) #status experimental <MAT>
F:45,47,62,119/Binding site: copper (His) #status predicted
F:56-145/Dsulfide bonds: #status predicted
F:62,70,79,82/Binding site: zinc (His, His, His, Asp) #status predicted
F:112/Active site: Arg #status predicted

Query Match          5.9%; Score 164; DB 2; Length 152;
Best Local Similarity 31.7%; Pred. No. 5e-05;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

QY      225 VHGTFDKGVGYDDELEVSYHLBGFNVSDHKDHLHDVOIYANGDLTSGCDNLGAKYDP-- 282
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       14 VKGIHFPGCGDPRPVTVTSVSGL-----KRLGHGFHVHALGSTTGCGNSTCPHNPNAS 67

```

```

OY 283 -----HEVDHSELGLDGLIHDDHGCVNESHRSKSWINIFEDSDSVGRSLAIH-QRDHLHK 336
Db 68 KENGAPEDEKRNHAGDGNVTAGADSVANINVTBDSQIPLTGPNSTIGRAVYVHADPDLGK 127

OY 337 -----SAKIACCVIG 346
Db 128 GGHELSKSTGNAAGRVACGIIG 149

RESULT 13
S22508
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) soda - rice
C:Species: Oryza sativa (rice)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 20-Jun-2000
C:Accession: S22508
R:Sakamoto, A.; Ohnuga, H.; Tanaka, K.
Plant Mol. Biol. 19, 323-327, 1992
A:Title: Nucleotide sequences of two cDNA clones encoding different Cu/Zn-superoxide dis-
A:Reference number: S22508; MUID:92322961; PMID:1623183
A:Accession: S22508
A:Molecule type: mRNA
A:Residues: 1-152 <SAK>
A:Cross-references: EMBL:D00999; NID:g218223; PTDN:BA00799.1; PID:g218224
A:Experimental source: clone RSODA
C:Genetics:
A:Gene: soda
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen
C:Superfamily: superoxide dismutase (Cu-Zn)
C:Keywords: copper; metalloprotein; oxidoreductase; zinc
F:45:47:62:119/Binding site: copper (His) #status predicted
F:56:145/disulfide bonds: #status predicted
F:62:70:79:82/Binding site: zinc (His, His, His, Asp) #status predicted
F:12/Active site: Arg #status predicted

Query Match 5.8%; Score 163; DB 2; Length 152;
Best Local Similarity 32.4%; Pred. No 5,9e-05;
Matches 46; Conservative 19; Mismatches 51; Indels 26; Gaps 4;

OY 225 VHGITIDFKQYGVGDLEVSYTHLEGFNVSDDKDHLHDVITANGDPLTSGCDNLGAKYDP-- 282
Db 14 VKGTHIFVQEDSDGTYVTGVSGL-----KFLGFIHIALGDTTGCMCKSTGPHNIPAG 67
OY 283 -----HEDYHSELGLDGLIHDDHGCVNESHRSKSWINIFGDSVLSGRSLAIH-QRDHLK 336
Db 68 KENGAPEDEKRNHAGDGNVTAGADSVANINVTBDSQIPLTGPNSTIGRAVYVHADPDLGK 127

OY 337 -----SAKIACCVIG 346
Db 128 GGHELSKSTGNAAGRVACGIIG 149

RESULT 14
S72235
superoxide dismutase (EC 1.15.1.1) (Cu-Zn) 4A, cytosolic [validated] - maize
C:Species: Zea mays (maize)
C:Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C:Accession: S72235; S78441; S07008; S20570
R:Kernodle, S.P.; Scandalios, J.G.
Genetics 144, 317-328, 1996
A:Title: A comparison of the structure and function of the highly homologous maize anti-
A:Reference number: S72234; MUID:97032947; PMID:8878695
A:Accession: S72235
A:Molecule type: DNA
A:Residues: 1-152 <KERN>
A:Cross-references: EMBL:U34727; NID:g1899026
R:Scandalios, J.G.
Submitted to the EMBL Data Library, August 1995
A:Reference number: S78441
A:Accession: S78441
A:Molecule type: DNA
A:Residues: 1-61, D, 63-152 <SCA>

```

A:Cross-references: EMBL:U34727; NID:g1890926; PID:AA49913.1; PID:g1885354  
R:Cannon, R.E.; Scandalios, J.G.  
Mol Gen Genet 219, 1-8, 1989

A:Title: Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase proteins in A:Reference number: S07007; MUID:90136495; PMID:2482436

A:Accession: S07008

A:Molecule type: mRNA

A:Residues: 1-61, D' 63A-152 <CAN>

A:Cross-references: EMBL:X17564; NID:g22484; PTDN:CAB57993.1; PID:g6018746

A:Experimental source: Indred Line W64A

A>Note: The authors translated the codon CAC for residue 61 as Pro and GAC for residu

A:Molecule type: protein

A:Accession: S20570

A:Residues: 2-20 <CAN2>

C:Genetics:

A:Gene: sod4A

A:Introns: 26/1; 60/1; 92/1; 102/3; 128/1; 146/1

C:Function:

A>Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg

C:Superfamily: superoxide dismutase (Cu-Zn)

C:Keywords: copper; cytosol; metalloprotein; oxidoreductase; zinc

F:2-152/Product: superoxide dismutase (Cu-Zn) #status experimental <MAT>

F:45,47,62,119/Binding site: copper (His) #status predicted

F:56-145/Dismutated bonds: #status predicted

F:62,70,79,82/Binding site: zinc (His, His, Asp) #status predicted

F:142/Active site: Arg #status predicted

Query Match 5.8%; Score 163; DB 2; Length 152;  
Best Local Similarity 31.7%; Pred. No. 5.9e-05;  
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

OY 225 VHGTRDKQGVGDLEVSYLEGFNVSDHDDHDLADVOIYANDLTSGCDNLGAKYPD-- 282  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 14 VKGIFFPTGGDGPRTAVTGSVGL-----KPGIGHGFHVHALDDTTNGCMSTGHNYNPAS 67

OY 283 -----HEDVYSLELDGDIHDHDCVNESHRYMINIFEDDSVLGRSLAH-QRDHLHK 336  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 68 KENGAPEDENRHAGDIGNTAGADGVANINVTDSQIPITGPSNISIGRAVVHADPDGLK 127

OY 337 -----SAKIACCVIIG 346  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 128 GGHLSKSTGNMGGRVACGIIIG 149

RESURF 15

JQ0915

Superoxide dismutase (EC 1.15.1.1) (Cu-Zn) - mouse

N:Alternate names: Proteasome chain C8

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1991 #sequence\_release 30-Sep-1991 #text\_change 20-Apr-2000

A:Accession: JQ0915; S00910; A37101

R:Benedetto, M.T.; Anzal, Y.; Gordon, J.W.

Gene 99, 191-195, 1991

A>Title: Isolation and analysis of the mouse genomic sequence encoding Cu2+-Zn2+ supe

A:Reference number: JQ0915; MUID:91216458; PMID:2022332

A:Accession: JQ0915

A:Molecule type: DNA

A:Residues: 1-154 <BEN>

A:Cross-references: GB:W60794

A>Note: The authors translated the codon CAT for residue 102 as Asp

A>Note: the sequence shown follows the authors' translation

R:Bewley, G.C.

Nucleic Acids Res. 16, 2728, 1988

A>Title: cDNA and deduced amino acid sequence of murine Cu-Zn superoxide dismutase.

A:Reference number: S00910; MUID:88203220; PMID:3362863

A:Accession: S00910

A:Molecule type: mRNA

A:Residues: 1-154 <BEM>

A:Cross-references: EMBE:X06683; NID:g54127; PIDN:CMA29880.1; PID:g54128

R:Plutcher, F.G.; Shmeer, M.; Esklinazi, D.; van der Gaag, H.; Huang, K.S.; Hulmes, J

J. Cell Biol. 111, 1217-1223, 1990

A>Title: Purification of an inhibitor of erythroid progenitor cell cycling and antago

A:Reference number: A37101; MUID:90361747; PMID:2391363  
 A:Accession: A37101  
 A:Molecule type: protein  
 A:Residues: 4-6, 'X', 8-23 <PLU>  
 C:Genetics:  
 A:Gene: SOD-1  
 A:Introns: 24/3; 57/1; 80/2; 119/3  
 C:Function:  
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen  
 C:Superfamily: superoxide dismutase (Cu,Zn)  
 C:Keywords: copper metalloprotein; oxidoreductase; zinc  
 F:47/49/64,121/Binding site: copper (His) #status predicted  
 F:58-147/Disulfide bonds: #status predicted  
 F:64,72,81,84/Binding site: zinc (His, His, Asp) #status predicted  
 F:144/Active site: Arg #status predicted

Query Match 5.8%; Score 162.5; DB 2; Length 154;  
 Best Local Similarity 30.1%; Pred. No. 6.5e-05;  
 Matches 44; Conservative 25; Mismatches 50; Indels 27; Gaps 6;

QY 225 VHGTFDKQVGYD-LEVSYLEGFNVSDDKDHLDVQIYANGDLTSGCDNLGAKYDPH 283  
 Db 15 VOGTHFEOKASGEFVLSGQITGL-TEGQHFVHQY-----CDNTQCTSGAPHPNPH 68  
 QY 284 EDYHS-----ELGDLGDIHDDHGVNESHRYSWINIFGDDSVLGRSIAIHQR-DHLH 335  
 Db 69 SKKHGPADEERHVGDLGVNAGKGVANVSIEDRVISLSEHSIIIGRTMVVHEKQDDLG 128  
 QY 336 K-----SAKIACCVTGRGO 349  
 Db 129 KCGNEESTKTNAGSRIACGVIGTINQ 154

Search completed: November 29, 2002, 12:36:07  
 Job time : 49 secs

THIS PAGE BLANK (USPTO)



Gencore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 10:12:53 ; Search time 26 Seconds  
(without alignments)  
792.836 Million cell updates/sec

Title: US-09-868-760-7  
Perfect score: 2794  
Sequence: 1 DGEQNDGQNKDHHDDHD.....MOGHEVESERVACCVIGRA 497

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 213   | 7.6         | 351    | 1  | HRPX_PLALO  |
| 2          | 178   | 6.4         | 852    | 1  | SRCH_RABIT  |
| 3          | 175   | 6.3         | 151    | 1  | SODC_HAIRO  |
| 4          | 173.5 | 6.2         | 151    | 1  | SODC_PEA    |
| 5          | 173.5 | 6.2         | 244    | 1  | SODE_RAT    |
| 6          | 171.5 | 6.1         | 251    | 1  | SODE_MOUSE  |
| 7          | 168   | 6.0         | 150    | 1  | SODE_MOUSE  |
| 8          | 167.5 | 6.0         | 244    | 1  | SODE_RABIT  |
| 9          | 167   | 6.0         | 153    | 1  | SODC_YEAST  |
| 10         | 164.5 | 5.9         | 153    | 1  | SODC_HORSE  |
| 11         | 164   | 5.9         | 151    | 1  | SOD4_MALIZ  |
| 12         | 164   | 5.9         | 151    | 1  | SOD5_MALIZ  |
| 13         | 163   | 5.8         | 151    | 1  | SOD1_ORYSA  |
| 14         | 163   | 5.8         | 152    | 1  | SODC_ZANAE  |
| 15         | 163   | 5.8         | 153    | 1  | SODC_CANAL  |
| 16         | 162.5 | 5.8         | 153    | 1  | SODC_MOUSE  |
| 17         | 161.5 | 5.8         | 153    | 1  | SODC_RAT    |
| 18         | 161   | 5.8         | 152    | 1  | SODC_ANACO  |
| 19         | 160.5 | 5.7         | 152    | 1  | SODC_CAVPO  |
| 20         | 160   | 5.7         | 151    | 1  | SODC_VITVI  |
| 21         | 159.5 | 5.7         | 151    | 1  | SOD1_BRAJU  |
| 22         | 159.5 | 5.7         | 151    | 1  | SRCH_HUMAN  |
| 23         | 159   | 5.7         | 166    | 1  | SODC_CARCA  |
| 24         | 158   | 5.7         | 151    | 1  | SODC_ORYSA  |
| 25         | 158   | 5.7         | 153    | 1  | SODC_SPILO  |
| 26         | 157.5 | 5.6         | 141    | 1  | SODC_PINSY  |
| 27         | 157.5 | 5.6         | 152    | 1  | SODC_RABIT  |
| 28         | 155.5 | 5.6         | 153    | 1  | SODC_HUMAN  |
| 29         | 154.5 | 5.5         | 153    | 1  | SODC_SCHWA  |
| 30         | 153.5 | 5.5         | 152    | 1  | SODC_SPILO  |
| 31         | 153   | 5.5         | 152    | 1  | SOD1_MESGR  |
| 32         | 153   | 5.5         | 154    | 1  | SODC_SCHPO  |
| 33         | 153   | 5.5         | 154    | 1  | SODC_SCHPO  |

|    |       |     |     |   |            |                     |
|----|-------|-----|-----|---|------------|---------------------|
| 34 | 152   | 5.4 | 151 | 1 | SODC_ARATH | P24704 arabidopsis  |
| 35 | 151.5 | 5.4 | 216 | 1 | SODE_ZANAE | O65175 zantedeschia |
| 36 | 151.5 | 5.4 | 220 | 1 | SODE_SOLCS | O04997 solidago ca  |
| 37 | 151   | 5.4 | 152 | 1 | SODE_SOLCS | O04996 solidago ca  |
| 38 | 151   | 5.4 | 201 | 1 | SODE_ONCVO | O07449 onchocerca   |
| 39 | 150   | 5.4 | 151 | 1 | SOD1_LYCES | P14830 lycopersico  |
| 40 | 150   | 5.4 | 152 | 1 | SODC_CARPA | O65768 carica papa  |
| 41 | 150   | 5.4 | 152 | 1 | SODC_CARPA | P11418 carica papa  |
| 42 | 150   | 5.4 | 217 | 1 | SODE_LYCES | P14831 lycopersico  |
| 43 | 148.5 | 5.3 | 151 | 1 | SODE_CEREL | O46412 cervus elap  |
| 44 | 148.5 | 5.3 | 154 | 1 | SODE_BRAE  | O73872 brachydanio  |
| 45 | 148.5 | 5.3 | 184 | 1 | SODE_CABEL | P34461 caenorhabdi  |

## ALIGNMENTS

| RESULT 1 | ID   | HRPX_PLALO  | STANDARD: | PRT: | 351 AA. |
|----------|--|---|-----------|------|---------|
| AC       | P04929:  |   |           |      |         |
| DT       | 13-AUG-1987 (Rel. 05, Created)   |   |           |      |         |
| DT       | 13-AUG-1987 (Rel. 05, Last sequence update)  |   |           |      |         |
| DT       | 15-JUL-1999 (Rel. 38, Last annotation update)  |   |           |      |         |
| DE       | Histidine-rich glycoprotein precursor.   |   |           |      |         |
| OS       | Plasmodium lophurae.   |   |           |      |         |
| OC       | Eukaryota, Alveolata, Apicomplexa, Haemosporidia, Plasmodium.  |   |           |      |         |
| OX       | NCBI_TaxID=5853;   |   |           |      |         |
| RN       | (1)  |   |           |      |         |
| RP       | SEQUENCE FROM N.A. Pubmed-6095114.   |   |           |      |         |
| RX       | MEDLINE=85061618; Feder R., Pavlovic A., Biobel G.,  |   |           |      |         |
| RA       | Pravich J.V., Feder R., Pavlovic A., Biobel G.,  |   |           |      |         |
| RT       | "Primary structure and genomic organization of the histidine-rich  |   |           |      |         |
| RL       | protein of the malaria parasite Plasmodium lophurae.";   |   |           |      |         |
| CC       | Nature 312:616-620(1984).  |   |           |      |         |
| CC       | -I- MISCELLANEOUS: IN THE INTRAERYTHROCYTIC STAGES OF DEVELOPMENT OF   |   |           |      |         |
| CC       | P. LOPHURAE IN DUCKS, THERE IS A SYNTHESIS OF A MAJOR PROTEIN THAT   |   |           |      |         |
| CC       | ACCUMULATES TO COMPRISE AT LEAST 50% OF THE CELLULAR MASS: THE   |   |           |      |         |
| CC       | HISTIDINE RICH PROTEIN.  |   |           |      |         |
| CC       | -----  |   |           |      |         |
| CC       | This SWISS-PROT entry is copyright. It is produced through a collaboration   |   |           |      |         |
| CC       | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |   |           |      |         |
| CC       | the European Bioinformatics Institute. There are no restrictions on its  |   |           |      |         |
| CC       | use by non-profit institutions as long as its content is in no way   |   |           |      |         |
| CC       | modified and this statement is not removed. Usage by and for commercial  |   |           |      |         |
| CC       | entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> |   |           |      |         |
| CC       | or send an email to <a href="mailto:license@sdb.ch">license@sdb.ch</a> ).  |   |           |      |         |
| CC       | -----  |   |           |      |         |
| DR       | EMBL: X01469; CAA25698.1; -  |   |           |      |         |
| DR       | PIR: A22692; KGZOH.  |   |           |      |         |
| KW       | Malaria; Repeat: Glycoprotein; Signal.   |   |           |      |         |
| FT       | CHAIN  | 24  | 47        |      |         |
| FT       | PROPEP   | 48  | 351       |      |         |
| FT       | SIGNAL   | 1   | 23        |      |         |
| FT       | CARBOHYD   | 59  | 90        |      |         |
| FT       | DOMAIN   | 59  | 90        |      |         |
| FT       | REPEAT   | 75  | 74        |      |         |
| FT       | REPEAT   | 59  | 90        |      |         |
| FT       | REPEAT   | 91  | 123       |      |         |
| FT       | DOMAIN   | 91  | 107       |      |         |
| FT       | REPEAT   | 108   | 123       |      |         |
| FT       | REPEAT   | 124   | 153       |      |         |
| FT       | DOMAIN   | 124   | 138       |      |         |
| FT       | REPEAT   | 139   | 153       |      |         |
| FT       | REPEAT   | 173   | 351       |      |         |
| FT       | DOMAIN   | 351   | 44032     |      |         |
| FT       | SEQUENCE   | 351 AA; 44032 MW; D19A48D47D890453 CPG64;                 |           |      |         |
| QY       | 12 DDHHDDHHDDHHDDDDDETHNVAO-CEMEPRNPMASLHHHHYSIELSKGCHAVYL 70  |   |           |      |         |
|          | Query Match  | 7.6%; Score 213; DB 1; Length 351;                        |           |      |         |
|          | Best Local Similarity  | 19.4%; Pred. No. 6.8e-09;                                 |           |      |         |
|          | Matches  | 83; Conservative 31; Mismatches 163; Indels 150; Gaps 17; |           |      |         |
|          | Matches  | 83; Conservative 31; Mismatches 163; Indels 150; Gaps 17; |           |      |         |

```
Db      67 EEEERRNEENNNRRNREERNRNEENNRRNNRN---NNNRPRNNNGN--NN    120
Oy      71 ELTLVGFNTSEDDNDNNHGLNLMLGMSAGSODSGELMAREKFAVDGSGLDVDDR   130
Db      121 NNNANNNNNNNNNNNNNNAANN-----NNEVNINN-----             152
Oy      131 GVAVENVNAVWLDIGSTARNTEALIGSMITLDSHPDADTSRYSLACSVTGAKARET   190
Db      153 ----NANNRPF-----NNNLSGVNNNARRNNNNNNNARRNN-----NNNNAAH-   195
Oy      191 AALLNELLEDRTTGAACSVASKYSPNOVKALNNYNGSTIDFKVGUGDLVEVLGEFNV   250
Db      196 -----NNNNNARRNNNN-----NNARRNNNNNNNNNNNNNNNNNNNNNNNGN-   245
Oy      251 SDNKKNLLDYUATYGNDLTSCGNISLKAKXDREDYUSLESDGLSDIDRDDNVESHRX   310
Db      246 --NNNNNNDAANNNN-----NNNNNNNNNN-----NDAN--NNNNN       282
Oy      311 SWNTGGDSLVLRSTAINDRDHLKSKAKTASCYLGSGNSRHETVARAKSVAPRTESG   370
Db      283 D-----AHHNNNNNDN-----NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN   295
Oy      371 LHNHGVSITFEOTREGSTNYMTADIKGFVSEDLSHNRGVOLHEMDMSGCHSLGRMY   430
Db      296 -NN-----NNNNNNNNNNNNNDANNNN--NNNDANNNNNNNNNDANNNN     337
Oy      431 NGHDAN 437
Db      338 NNNNDAN 344

RESULT 2
SRCH_RABIT ID SRCH_RABIT STANDARD: PRT; 852 AA.

AC PI6230;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sarcoplasmic reticulum histidine-rich calcium-binding protein
DE precursor.
DE HRC OR MCP.
GN Oryctolagus cuniculus (Rabbit).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RA [1]
RA SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA TISSUE=Skeletal muscle;
RX MEDLINE=90036884; PubMed=2808365;
RA Holman S.L., Goldstein J.L., Orth K., Moosm C.R., Slaughter C.A.,
RA Brown M.S.;
RT "Molecular cloning of a histidine-rich Ca2+-binding protein of
RT sarcoplasmic reticulum that contains highly conserved repeated
RT elements."
RL J.Biol.Chem. 264:18083-18090(1989).
CC -I- FUNCTION: MAY PLAY A ROLE IN THE REGULATION OF CALCIUM
CC SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC MUSCLE.
CC -I- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.
CC -I- SIMILARITY: STRONG, TO HUMAN HRC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost. On
CC its European Bioinformatics Institute, there are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to licenses@sdb.ch).

DR EMBL; J05080; AAA31279.1; -.
DR PIR; A34373; A34373.
DR InterPro; IPR002134; HCP_10.
DR PROSITE; PS00328; HCP_10.
KW Calcium-binding; Signal; Repeat.
```

[illegible]

DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyrosidae; Halocynthia.  
OX NCBI\_TaxID=7729;  
RN (1)  
RP SEQUENCE, AND CHARACTERIZATION.  
RC TISSUE=Hemocyte, and Plasma;  
RA MEDLINE=99302489, Pubmed=103742259;  
AB Abe Y., Ishikawa G., Satoh H., Azumi K., Yokosawa H.;  
RT Primary structure and function of superoxide dismutase from the  
RT ascidian Halocynthia roretzi.  
RL Comp. Biochem. Physiol. 122B:321-326(1999).  
CC -1- FUNCTION: Destroys radicals which are normally produced within the  
CC cells and which are toxic to biological systems. The plasma  
CC superoxide dismutase has phagocytosis-stimulating activity and may  
CC play an important role in the biological defenses of the organism.  
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -1- COFACTOR: Copper and zinc (By similarity).  
CC -1- ENZYME REGULATION: INHIBITED BY KCN AND DIETHYLDITHIOCARBAMATE.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- PTM: ACETYLATED N-TERMINAL SERINE.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
DR HSSP: P15107: 1XSO.  
DR InterPro: IPR001424: SOD\_CU\_ZN.  
DR Pfam: PF00080: sodcu.1.  
DR PRINTS: PR00068: CUZNDISMUTASE.  
DR ProDom: PD000469: SOD\_CU\_ZN.1.  
DR PROSITE: PS00087: SOD\_CU\_ZN.1; 1.  
DR PROSITE: PS00332: SOD\_CU\_ZN.2; FALSE\_NEG.  
KM Oxidoreductase; Copper; zinc.  
FT METAL 43 43 COPPER (BY SIMILARITY).  
FT METAL 45 45 COPPER AND ZINC (BY SIMILARITY).  
FT METAL 60 60 COPPER AND ZINC (BY SIMILARITY).  
FT METAL 68 68 ZINC (BY SIMILARITY).  
FT METAL 77 77 ZINC (BY SIMILARITY).  
FT METAL 80 80 ZINC (BY SIMILARITY).  
FT METAL 117 117 COPPER (BY SIMILARITY).  
FT METAL 143 143 BY SIMILARITY.  
SQ DISULFID 151 AA; 15489 MW; 162F181A82275F0 CRC64;  
Query Match 6.3%; Score 175; DB 1; Length 151;  
Best Local Similarity 29.6%; Pred. No. 1.6e-06;  
Matches 42; Conservative 26; Mismatches 48; Indels 26; Gaps 5;  
QY 225 VHCIDFKQVGYGDELEVSYLEGFNVSDDKHDLHDVQIYANGDLISGCDNLGAKYDPHE 284  
DB 12 VKETLNFKODALIGCTVGEVSGL-IPKGHGFHIEY-----GDLNFGCTSSGCHFNPK 65  
QY 285 DYHS-----ELGDLGDIHDDHGVNESHRYSMINIFGDDSVLGRSIAHQ-RDHL-- 334  
DB 66 QIHGAREDDIRHVGDLGNTADSSGAVNITDMISLSTGEHSIIIGRAVVVHAGEDDLGK 125  
QY 335 -----HKSAKIACVIG 346  
DB 126 GGHEDSKTGHAGRLSGVIG 147  
RESULT 4  
SODC\_PEA  
ID SODC\_PEA STANDARD: PRT: 151 AA.  
AC 002610:  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1999 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
GN SODC.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
OX NCBI\_TaxID=3888;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA White D.A., Zillinskas B.A.;  
RT "Nucleotide sequence of a complementary DNA encoding pea cytosolic  
RT copper/zinc superoxide dismutase";  
RL Plant Physiol. 96:1391-1392(1991).  
CC -1- FUNCTION: Destroys radicals which are normally produced within the  
CC cells and which are toxic to biological systems.  
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -1- COFACTOR: Copper and zinc.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: M63003: AAA3659.1; -.  
DR HSSP: P07505: 1SRD.  
DR InterPro: IPR001424: SOD\_CU\_ZN.  
DR Pfam: PF00080: sodcu.1.  
DR PRINTS: PR00068: CUZNDISMUTASE.  
DR ProDom: PD000469: SOD\_CU\_ZN.1.  
DR PROSITE: PS00087: SOD\_CU\_ZN.1; 1.  
DR PROSITE: PS00332: SOD\_CU\_ZN.2; 1.  
KM Oxidoreductase; Copper; zinc.  
FT INITMET 0 0 BY SIMILARITY.  
FT METAL 44 44 COPPER (BY SIMILARITY).  
FT METAL 46 46 COPPER (BY SIMILARITY).  
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).  
FT METAL 69 69 ZINC (BY SIMILARITY).  
FT METAL 78 78 ZINC (BY SIMILARITY).  
FT METAL 81 81 ZINC (BY SIMILARITY).  
FT METAL 118 118 COPPER (BY SIMILARITY).  
FT METAL 144 144 BY SIMILARITY.  
SQ DISULFID 151 AA; 15192 MW; 9BCFA53DD2CB1719 CRC64;  
Query Match 6.2%; Score 173.5; DB 1; Length 151;  
Best Local Similarity 31.4%; Pred. No. 2.1e-06;  
Matches 49; Conservative 26; Mismatches 48; Indels 33; Gaps 5;  
QY 211 VRSNTQPKALHNHAGCTIDFKQVGYGDELEVSYLEGFNVSDDKHDLHDVQIYANGDLT 270  
DB 6 VLSNSNE-----VSGTINFSQEGNGPTVYGTLAGL-----KFLGFIHIALGDTT 52  
QY 271 SGCDNLGAKYDPH-----EDYHSELGDLGDIHDDHGVNESHRYSMINIFGDDSVLG 323  
DB 53 NGCISIGPHFNNGKHEGAPDEDETRHAGDLGNINVGDDGVSTITDNIHPLTGINSIIG 112  
QY 324 RSFAIH-ORDHLK-----SAKIACVIG 346  
DB 113 RAVVVAADPDDLKGKGHLSKTGTGNAGGRVACVIG 148  
RESULT 5  
SODE\_RAT  
ID SODE\_RAT STANDARD: PRT: 244 AA.  
AC 008420: Q64667;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)  
DE (EC-SOD) (Superoxide dismutase B).  
GN SOD3 OR SOD-3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic;  
 RX MEDLINE=93319511; PubMed=8328962;  
 RA Perry A.C.F., Jones R., Hall L.;  
 RT "Isolation and characterization of a rat cDNA clone encoding a  
 RT secreted superoxide dismutase reveals the epidiolins to be a major  
 RT site of its expression."  
 RL Biochem. J. 293:21-25(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=94043314; PubMed=8227019;  
 RA Williams J., Zwijsen A., Slegers H., Nicolai S., Bettadapura J.,  
 RA Raymakers J., Scarce T.;  
 RT "Purification and sequence of rat extracellular superoxide dismutase  
 RT B secreted by C6 glioma."  
 RL J. Biol. Chem. 268:24614-24621(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC STRAIN=Sprague-Dawley; TISSUE=Uterus;  
 RX MEDLINE=96224263; PubMed=8643556;  
 RA Carlson L.M., Marklund S.M., Edlund T.;  
 RT "The rat extracellular superoxide dismutase dimer is converted to a  
 RT tetramer by the exchange of a single amino acid."  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:5219-5222(1996).  
 CC -1- FUNCTION: Destroys radicals which are normally produced within the  
 CC cells and which are toxic to biological systems.  
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) -> O(2) + H(2)O(2).  
 CC -1- COPROCTOR: Copper and zinc.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).  
 CC -----  
 DR EMBL: X68041; CAA48177.1; -;  
 DR EMBL: 224721; CAA80849.1; -;  
 DR EMBL: X94371; CAA64149.1; -;  
 DR PIR: S34055; S34055.  
 DR HSSP: P00441; LSOS.  
 DR InterPro: IPR001424; SOD\_CU\_ZN.  
 DR Pfam: PF000080; sdcu.1.  
 DR PRINTS: PD00068; CUZNDISMUTASE.  
 DR PRODOM: PD000469; SOD\_CU\_ZN.1.  
 DR PROSITE: PS00087; SOD\_CU\_ZN.1.  
 DR PROSITE: PS00332; SOD\_CU\_ZN.2; 1.  
 KM Oxidoreductase; Copper; Zinc; Glycoprotein; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 244  
 FT METAL 121 121  
 FT METAL 123 123  
 FT METAL 138 138  
 FT METAL 146 146  
 FT METAL 149 149  
 FT METAL 152 152  
 FT METAL 188 188  
 FT METAL 214 214  
 FT DISULFID 132 214  
 FT CABOHAD 114 114  
 FT MURAGEN 48 48  
 FT CONFLICT 235 235  
 FT CONFLICT 237 237  
 SO SEQUENCE 244 AA; 26620 MW; B66726501CE58614 CXC64;  
 Query Match 6.2%; Score 173.5; DB 1; Length 244;

Best Local Similarity 29.7%; Pred. No. 3.8e-06;  
 Matches 54; Conservative 28; Mismatches 61; Indels 39; Gaps 10;  
 QY 194 LHHELEDKTEHNAACDYRSNTQPKAL-----HHNVKGTDFKQVGYGD-LEVSXHEGF 248  
 DB 55 LKQREADARMAHVCV-----QPSAMLPPOQPTGTGLVLRQLGSPSRLEASFNLGEF 109  
 QY 249 NVSDHDKDLHDVOITANGDLSGCDNLCAKYPDEHDYHSELGLGDIHDDHG--VYNE 306  
 DB 110 PAEONTSNHA--IHVEFGDLSGCESTSPHNP-----LGVPHDPHDEDFEYVRD 160  
 QY 307 ---SHRYSW-INIFGDDSVLGRSIAIHQ-RDHLKSA-----KIACVYGRG 348  
 DB 161 GLWKHRHGLATSLAGPHSTLGRAVYVHAGEDDLKGGNQASVONGNAGRRLACCVGTG 220  
 QY 349 QS 350  
 DB 221 NS 222  
 RESULT 6  
 SODE\_MOUSE  
 ID SODE\_MOUSE STANDARD; PRI: 251 AA.  
 AC 009164;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)  
 DE (EC-SOD).  
 GN SOD3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-36.  
 RC STRAIN=C57BL/6 J CBA; TISSUE=Lung;  
 RX MEDLINE=98026284; PubMed=9376114;  
 RA Foltz R.J., Guan J., Seidlin M.F., Oury T.D., Englund J.J., Crapo J.D.;  
 RT "Mouse extracellular superoxide dismutase: primary structure, tissue-  
 RT specific gene expression, chromosomal localization, and lung in situ  
 RT hybridization."  
 RL Am. J. Respir. Cell Mol. Biol. 17:393-403(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=97306432; PubMed=9163733;  
 RA Suh Y.-S., Wada K., Yamashita T., Kikuchi T., Foltz R.J., Tanaka K.,  
 RA "Sequence analysis, tissue expression and chromosomal localization of  
 RT a mouse secreted superoxide dismutase gene."  
 RL Mol. Cells 7:204-207(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RA Siegfried M.R., Schultz D., Harrison D.G., Fukui T.;  
 RT "Murine extracellular superoxide dismutase genomic sequence."  
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Destroys radicals which are normally produced within the  
 CC cells and which are toxic to biological systems.  
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) -> O(2) + H(2)O(2).  
 CC -1- COPROCTOR: Copper and zinc (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).  
 CC -----

```

DR EMBL: U38261; AAB51106.1; -
DR EMBL: D50856; BAA23493.1; -
DR EMBL: AF223251; AAF27932.1; -
DR HSSP: P00445; 1JCV.
DR MGD: MGI:103181; SOD3.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
KW Oxidoreductase; Copper; Zinc; Glycoprotein; Signal.
FT SIGNAL 1 15
FT PROPEP 16 24
FT CHAIN 25 251
FT METAL 128 128 EXTRACELLULAR SUPEROXIDE DISMUTASE [CU-
FT METAL 130 130 ZN].
FT METAL 145 145 COPPER (BY SIMILARITY).
FT METAL 153 153 COPPER AND ZINC (BY SIMILARITY).
FT METAL 156 156 ZINC (BY SIMILARITY).
FT METAL 159 159 ZINC (BY SIMILARITY).
FT METAL 195 195 ZINC (BY SIMILARITY).
FT METAL 195 195 COPPER (BY SIMILARITY).
FT DISULFID 121 121 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 251 AA; 27392 MW; 9EAF850E458966C4 CRC64;

Query Match 6.1%; Score 171.5; DB 1; Length 251;
Best Local Similarity 30.6%; Pred. No. 5.6e-06;
Matches 55; Conservative 23; Mismatches 57; Indels 45; Gaps 11;

OY 199 EEKRTGTHVHNCVRSNTIOPKAL---HHVHGTFIDKQVGYD-LEVSYLEGFAVSD 253
DB 67 EYDAAMAHMICRV-----QPSATLPPDPOQITGLVLEFRLQGPSRLAEVFSLEGFABEN 121
OY 254 HKD---HLHDVQIYANGTLSCDNIGKXPDHEDVSELDGDIHDDHG--VYNES- 307
DB 122 ASRAIAHVEF-----GDLSCGCBSTGTHVPMVPPQ-----HPPDFGFFVVRNQ 169
OY 308 ---HRS-WINIFGDSVLSGSIATHQ-RDLHLMSA-----KACCVIRGQS 350
DB 170 LMRHRYGLTASLAGFRAHILGHSVYHAGEDLDGKGQASLQNGNRRLACCVGTSSS 229

RESULT 7
SODE_MAIZE STANDARD; PRT; 150 AA.
ID SODE_MAIZE
AC P11428;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 2 (EC 1.15.1.1).
GN SODCC.1 OR SOD2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_Taxid=4577;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=87092359; PubMed=3467349;
RA Cannon R.E., White J.A., Scandalios J.G.;
RT "Cloning of cDNA for maize superoxide dismutase 2 (SOD2).";
RL Proc. Natl. Acad. Sci. U.S.A. 84:179-183(1987).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=87249683; PubMed=3597043;
RA Cannon R.E., Scandalios J.G.;
RT "The superoxide dismutase-2 gene of maize.";
RL Isozymes Curr. Top. Biol. Med. Res. 14:73-81(1987).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

```

```

CC -1- CORAFCTOR: Copper and zinc (by similarity).
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M15175; AAA3511.1; -
DR EMBL: M54936; AAA3510.1; -
DR PIR: A29077; A29077.
DR HSSP: P07505; 1SRD.
DR MaizeDB: 47586; -.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
KW Oxidoreductase; Copper; Zinc; Multigene family.
FT INIT MET 0 0
FT METAL 43 43 COPPER (BY SIMILARITY).
FT METAL 45 45 COPPER (BY SIMILARITY).
FT METAL 60 60 COPPER AND ZINC (BY SIMILARITY).
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 77 77 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 117 117 COPPER (BY SIMILARITY).
FT DISULFID 54 143 BY SIMILARITY.
SQ SEQUENCE 150 AA; 14973 MW; C63A117072C37873 CRC64;

Query Match 6.0%; Score 168; DB 1; Length 150;
Best Local Similarity 31.9%; Pred. No. 5.3e-06;
Matches 46; Conservative 22; Mismatches 46; Indels 30; Gaps 5;

OY 225 VHGTFIDKQVGYDLEVSYLEGFAVSDHDKHLHDVQIYANGTLSCDNIGAKYD-- 282
DB 12 VAGTIFFSOEGDGPPTVVGSTISGL-----KPGDGHVHALDQDTNGSMSTGPHFNPVG 65
OY 283 -----HEDVHSELGDIHDDHGVNESHRSYWINIFGDSVLSGSIATHQ-RDHL 334
DB 66 KEHGAPEDEDRHA--GDLGNTVAGEDGVVNVNITDSDIPLAGHSITIGRAVYVHADPDL 123
OY 335 HK-----SAKIAVCVIG 346
DB 124 GKGGHLSKSTGNACGRVACGII 147

RESULT 8
SODE_RABBIT STANDARD; PRT; 244 AA.
ID SODE_RABBIT
AC P41975;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
GN (EC-SOD).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white;
RA Laukkanen M.O., Aittomaki S.J., Hiltunen T.P., Yla-Herttuala S.;
RL submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN 12

```

RP SEQUENCE FROM N.A.  
 RC STRAIN-New Zealand white; TISSUE-Heart;  
 RA Laakkonen M.O., Hiltunen M.O., Alttomeki S., Janne J.,  
 RA Viala-Herttuala S.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Laakkonen M.O., Alttomeki S., Mannermaa S., Hiltunen M.O.,  
 RA Viala-Herttuala S.;  
 RT Cloning and characterization of rabbit extracellular superoxide  
 RT dismutase.\*;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 43-155 FROM N.A.  
 RC STRAIN-New Zealand white; TISSUE-Aorta;  
 RA Hiltunen T.P., Nikkari T., Viala-Herttuala S.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Destroys radicals which are normally produced within the  
 CC cells and which are toxic to biological systems.  
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
 CC -1- COFACTOR: Copper and zinc (By similarity).  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CO-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE CO-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z67878; CAA91785.1;  
 DR EMBL: Y13339; CAA73783.1;  
 DR EMBL: AJ007044; CAA07431.1;  
 DR EMBL: X78139; CAA55018.1;  
 DR HSSP: P00442; ICBJ.  
 DR InterPro: IPR001424; SOD\_CU\_ZN.  
 DR Pfam: PF00080; SODCU.1.  
 DR PRINTS: PR00068; CUZNDISMUTASE.  
 DR PRODOM: PD000469; SOD\_CU\_ZN.1.  
 DR PROSITE: PS00087; SOD\_CU\_ZN.1;  
 DR PROSITE: PS00332; SOD\_CU\_ZN.2;  
 KW Oxidoreductase; Copper; zinc; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 1 19  
 FT METAL 118 118  
 FT METAL 120 120  
 FT METAL 135 135  
 FT METAL 143 143  
 FT METAL 146 146  
 FT METAL 149 149  
 FT METAL 185 185  
 FT DISULFID 129 211  
 FT CARBOHYD 111 111  
 FT CONFLICT 53 53  
 FT SEQUENCE 244 AA; 25688 MW; 7C9B1C5994FE25 CRC64;  
 Query Match 5.0%; Score 167.5; DB 1; Length 244;  
 Best Local Similarity 29.2%; Pred. NO. 1.1e-05;  
 Matches 50; Conservative 19; Mismatches 67; Indels 35; Gaps 7;  
 Oy 349 QSPHEIYHAKVCYVRPNTSTGLHHVSGSTFEOT-PEGSTHMTADLKGFNNSSEDLSIH 407  
 Db 56 QGEFAGALHVAICRVPASATLDAQPRVSLVFRQLGQADLEAFPLEGFPVEANLS-- 113  
 Oy 408 RHGVLHMGDMSHGCHSLGRMYHGHDAHDKRRDGLGV-----ID 450  
 Db 114 SRAIHVHQFDLQGGCGSGAHNPLAVQH-PQHGDEGNGFAVNRGLMKYRSLAASLA 172  
 Oy 451 DSHGIVSTFTFDHLNVEDLNARSLVIMOGHEVESE-----RVACCVIG 495

Db 173 GPHSIV-GRAYVYHAGEDDLG-----RGGNASVENGNGNAGRFLACCVYG 215  
 RESULT 9  
 ID SODC\_YEAST STANDARD; PRT; 153 AA.  
 AC P00445;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
 GN SOD1 OR YJR104C OR J1968.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88263032; PubMed=3290902;  
 RA Berningham-McDonogh O., Gralla E., Valentine J.;  
 RT "The copper, zinc-superoxide dismutase gene of *Saccharomyces*  
 RT *cerevisiae*: cloning, sequencing, and biological activity.\*";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:4789-4793(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Remazan Red M., Kirchath L., Hollenberg C.P.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE.  
 RA Johansen J.T., Overballe-Petersen C., Martin B., Hasemann V.,  
 RA Svendsen I.;  
 RT "The complete amino acid sequence of copper, zinc superoxide dismutase  
 RT from *Saccharomyces cerevisiae*.";  
 RL Carlsberg Res. Commun. 44:201-217(1979).  
 RN [4]  
 RP SEQUENCE.  
 RX MEDLINE=80227835; PubMed=6993479;  
 RA Stehman H.M.;  
 RT "The amino acid sequence of copper-zinc superoxide dismutase from  
 RT bakers' yeast.\*";  
 RL J. Biol. Chem. 255:6758-6765(1980).  
 RN [5]  
 RP SEQUENCE OF 1-10.  
 RC STRAIN-X2180-1A;  
 RA Fruhiger S., Hughes G.J., Sanchez J.-C., Hochstrasser D.F.;  
 RL Submitted (FEB-1996) to the SWISS-PROT data bank.  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=92126276; PubMed=1772629;  
 RA Djionovic K., Gatti G., Coda A., Antolini L., Pelosi G., Desideri A.,  
 RA Falconi M., Marmocchi F., Rotilio G., Bolognesi M.;  
 RT "Structure solution and molecular dynamics refinement of the yeast  
 RT Cu,Zn enzyme superoxide dismutase.\*";  
 RL Acta Crystallogr. B 47:918-927(1991).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=92292161; PubMed=1602482;  
 RA Djionovic K., Gatti G., Coda A., Antolini L., Pelosi G., Desideri A.,  
 RA Falconi M., Marmocchi F., Rotilio G., Bolognesi M.;  
 RT "Crystal structure of yeast Cu,Zn superoxide dismutase.  
 RT Crystallographic refinement at 2.5-A resolution.\*";  
 RL J. Mol. Biol. 225:791-809(1992).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).  
 RX MEDLINE=99152006; PubMed=10026301;  
 RA Hart P.D., Baldwin M.H., Ogihara N.L., Nerstissian A.M., Weiss M.S.,  
 RA Valentine J.S., Eisenberg D.;  
 RT "A structure-based mechanism for copper-zinc superoxide dismutase.\*";  
 RL Biochemistry 38:2167-2178(1999).  
 CC -1- FUNCTION: Destroys radicals which are normally produced within the  
 CC cells and which are toxic to biological systems.  
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).

CC -1- COFACTOR: Copper and zinc.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J03279; AAC34543.1; -;  
DR EMBL: Z49604; CA89634.1; -;  
DR PIR: A36171; DSHYC.  
DR PDB: 1SDY; 31-JAN-94.  
DR PDB: 1YSO; 10-JUN-96.  
DR PDB: 1JCV; 08-MAR-96.  
DR PDB: 2JCW; 08-JUN-99.  
DR PDB: 1B4L; 23-DEC-99.  
DR PDB: 1B4T; 23-DEC-99.  
DR SWISS-2DPAGE: P00445; YEAST.  
DR SGI: S003865; SOD1.  
DR InterPro: IPR001424; SOD\_CU\_ZN.  
DR Pfam: PF00080; sodcu.1.  
DR PRINTS: PR00068; CUZNDISMASE.  
DR PRODOM: PD000469; SOD\_CU\_ZN.1.  
DR PROSITE: PS00087; SOD\_CU\_ZN.1; 1.  
DR PROSITE: PS00332; SOD\_CU\_ZN.2; 1.  
KW Oxidoreductase; Copper; Zinc; 3D-structure.  
FT INIT MET 0  
FT METAL 46  
FT METAL 46  
FT METAL 63  
FT METAL 63  
FT METAL 71  
FT METAL 71  
FT METAL 80  
FT METAL 80  
FT METAL 83  
FT METAL 83  
FT METAL 120  
FT METAL 120  
FT DISULFID 57  
FT DISULFID 146  
FT CONFLICT 55  
FT CONFLICT 55  
FT CONFLICT 92  
FT STRAND 2  
FT STRAND 8  
FT STRAND 14  
FT STRAND 20  
FT STRAND 28  
FT STRAND 35  
FT STRAND 41  
FT STRAND 48  
FT STRAND 54  
FT STRAND 57  
FT STRAND 58  
FT STRAND 60  
FT STRAND 63  
FT STRAND 67  
FT STRAND 74  
FT STRAND 75  
FT STRAND 81  
FT STRAND 82  
FT STRAND 83  
FT STRAND 89  
FT STRAND 91  
FT STRAND 92  
FT STRAND 95  
FT STRAND 101  
FT STRAND 105  
FT STRAND 109  
FT STRAND 110  
FT STRAND 111  
FT STRAND 111  
FT STRAND 113  
FT STRAND 114  
FT STRAND 116  
FT STRAND 119  
FT STRAND 126  
FT STRAND 127  
FT STRAND 133  
FT STRAND 133  
FT STRAND 134  
FT STRAND 137  
FT STRAND 138  
FT STRAND 139  
FT STRAND 148  
FT STRAND 150  
FT STRAND 151  
SQ SEQUENCE 153 AA; 15723 MW; 4B431A9B5D3211BE CRC64;  
Query Match 6.0%; Score 167; DB 1; Length 153;  
Best Local Similarity 30.1%; Pred. No. 6.4e-06;  
Matches 43; Conservative 26; Mismatches 48; Indels 26; Gaps 5;  
OY 225 VHGTFIDKQVGVGD-LEVSYHLSEGFNVSDDHKHLHDVQYANGDLTSGCDNLGAKYDPH 283

DB 13 VSGVKKFEQASESEPTYSYEIAGNSPNAERGFHIEF-----GATNGCVSAGHFHMF 67  
OY 284 EBYHS-----ELGDDGDIHDDHGVNESHRYSWINIFGDSVGSIAHQ-RDLH 335  
DB 68 KTHKAPFDEVRHGVKMDKNTDENGVAKGSFKDSLKLIGPYSVGRSVYIHAGDDLG 127  
OY 336 K-----SAKIACCYIG 346  
DB 128 KQTEBSLKTGNAGPRACGVIG 150  
RESULT 10  
SODC\_HORSE STANDARD; PRT; 153 AA.  
ID SODC\_HORSE  
AC P00443;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
GN SOD1.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97080551; PubMed=8921896;  
RA de la Rúa-Domenech R., Wiedmann M., Mohammed H.O., Cummings J.F.,  
RA Divers T.J., Batt C.A.;  
RT "Equine motor neuron disease is not linked to Cu/Zn superoxide  
RT dismutase mutations: sequence analysis of the equine Cu/Zn superoxide  
RT dismutase cDNA.";  
RT Gene 178:83-88(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=99261591; PubMed=10331206;  
RA Ishida N., Katayama Y., Sato F., Hasegawa T., Mukoyama H.;  
RT "The cDNA sequences of equine antioxidative enzyme genes Cu/Zn-SOD and  
RT Mn-SOD, and these expressions in equine tissues.";  
RT J. Vet. Med. Sci. 61:291-294(1999).  
RN [3]  
RP SEQUENCE.  
RX MEDLINE=82052979; PubMed=7298616;  
RA Lerch K., Ammer D.;  
RT "Amino acid sequence of copper-zinc superoxide dismutase from horse  
RT liver.";  
RT J. Biol. Chem. 256:11545-11551(1981).  
CC -1- FUNCTION: Destroys radicals which are normally produced within the  
CC cells and which are toxic to biological systems  
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) -> O(2) + H(2)O(2).  
CC -1- COFACTOR: Copper and zinc.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U38956; AAC48682.1; -;  
DR EMBL: AB001692; BAA76921.1; -;  
DR PIR: A00515; DSHOC2.  
DR HSSP: P00442; 1CBJ.  
DR InterPro: IPR001424; SOD\_CU\_ZN.  
DR Pfam: PF00080; sodcu.1.  
DR PRINTS: PR00068; CUZNDISMASE.  
DR PRODOM: PD000469; SOD\_CU\_ZN.1.

```

DR PROSITE: PS00087; SOD_CU_ZN.1: 1
DR PROSITE: PS00332; SOD_CU_ZN.2: 1
KW Oxidoreductase; Copper; Zinc; Acetylation.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION.
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 48 48 COPPER (BY SIMILARITY).
FT METAL 63 63 COPPER AND ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 80 80 ZINC (BY SIMILARITY).
FT METAL 83 83 ZINC (BY SIMILARITY).
FT METAL 120 120 COPPER (BY SIMILARITY).
FT DISULFID 57 146 BY SIMILARITY.
SQ SEQUENCE 153 AA; 15940 MW; 202A190DEFA043 CRC64;

Query Match 5.9%; Score 164.5; DB 1; Length 153;
Best Local Similarity 31.0%; Pred. No. 9.9e-06;
Matches 45; Conservative 26; Mismatches 43; Indels 31; Gaps 6;

QY 225 VHGTFKQYGVGDLEVSYLEGF--NVSDDHKDHLHDVOIYANGDLTSGCDNLGAKYD 281
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DQ 14 VHGVIHEEDQDEGGPRV---LKGFTGELTKGDHGRHWF---GNTRGCTTAADAHN 65
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 282 P-----HEDYISELSDGDIHDHDCGVESHVSHRYSGINIFGDDVSLGSIQHQR-DH 333
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DQ 66 PLSKKHGGRPDEHRHNGDLDGNTAENCKADKVDKMSYISLGRKHSIIIGRTMYVHEKODD 125
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 334 LHK-----SAKIACVCIG 346
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DQ 126 LKGKNEESTKTGNAGSRACGIV 150
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 11
SOD4_MAIZE STANDARD; PRT: 151 AA.
ID SOD4_MAIZE STANDARD; PRT: 151 AA.
AC P23345;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 4A (EC 1.15.1.1).
GN SODCC.3 OR SOD4A.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Palaeodae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RX MEDLINE=90136495; PubMed=2482436;
RA Cannon R.E.; Scandallos J.G.;
RT "Two CDNs encode two nearly identical Cu/Zn superoxide dismutase
   proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: COPPER and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X17564; -. NOT_ANNOTATED_CDS.
DR PIR: S07007; S07007.
DR HSSP: P07505; ISR0.
DR MaizADB: 47586; -.

```

DR InterPro:IPR001424; sod\_cu\_zn.  
DR Pfam:PF00080; sodcu.1.  
DR PRINTS:PR00068; CUZNDISMTASE.  
DR ProDom:PD000469; SOD\_CU\_ZN.1.  
DR PROSITE:PS00087; SOD\_CU\_ZN.1.1.  
DR PROSITE:PS00332; SOD\_CU\_ZN.2.1.  
KW Oxidoreductase; Copper; Zinc; Multigene family.  
FT INIT\_MET 0 0  
FT METAL 44 44 COPPER (BY SIMILARITY).  
FT METAL 46 46 COPPER (BY SIMILARITY).  
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).  
FT METAL 69 69 ZINC (BY SIMILARITY).  
FT METAL 78 78 ZINC (BY SIMILARITY).  
FT METAL 81 81 ZINC (BY SIMILARITY).  
FT METAL 118 118 ZINC (BY SIMILARITY).  
FT DISULFID 55 144 COPPER (BY SIMILARITY).  
SQ SEQUENCE 151 AA; 14983 MW; 965226F86C919E58 CRC64;  
Query Match 5.9%; Score 164; DB 1; Length 151;  
Best Local Similarity 31.7%; Pred. No. 1.1e-05;  
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;  
QY 225 VHGTFPKQVGYGDELVSYHLEGFNVSDHDDHDLHDVQIYANGDLTSGCDNLGAKYP-- 282  
Db 13 VKGTFEFGDEGDPPTVTVGVSGL-----PRGHLGHVHALDDTTNGCSTGPHNPAS 66  
QY 283 -----HEDHSELGDLGDIHDDHGVNESHRYEMINIFGDDSYLGRSLAH-ORDHLRK 336  
Db 67 KKHGAPEDEENRHHGDTGNVTAGADCVANINVTDSQIFLTPGNSITGRVAVVHADPPDLQK 126  
QY 337 -----SAKIACCVIG 346  
Db 127 GGHBLSKSTGNAGRVACGIIG 148  
RESULT 12  
SOD5\_MAIZE  
ID SOD5\_MAIZE STANDARD; PRT; 151 AA.  
AC P23346;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] 4AP (EC 1.15.1.1).  
GN SODCC.2 OR SOD4P.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Focaceae; PACC clade;  
OC Panicoideae; Andropogoneae; Zea.  
OX NCBI\_Taxid:4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-90136495; PubMed-2482436;  
RA Cannon R.E., Scandalios J.G.;  
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase  
RT proteins in maize."  
RL Mol. Gen. Genet. 219:1-8(1989).  
CC -!- FUNCTION: Destroys radicals which are normally produced within the  
CC cells and which are toxic to biological systems.  
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H2(O)2.  
CC -!- COFACTOR: Copper and zinc.  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X17565; CAB57992.1; ALT\_SEQ.



```

DR PIR: S07008; S07008.
DR HSSP: P07505; 1SRD.
DR MaizeDB: 47586; -.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
DR Oxidoreductase; Copper; Zinc; Multigene family.
FW INIT_MET 0 0
FT METAL 44 46 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT DISULFID 55 144 COPPER (BY SIMILARITY).
SQ SEQUENCE 151 AA; 14939 MW; 9C7E572A6C1AEFD CRC64;

Query Match
Best Local Similarity 31.7%; Pred. No. 1.1e-05; Length 151;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

QY 225 VHGTFKQVGYGDLGVSYHLEGFNVSDHKDHLVDQIYANGDITLSCDNLGAKYDP-- 282
DB 13 VKGTIFFTQEGDGPVAVTGSVGL-----KPLGHGFHVALDITNGCMSTGPHNPNAS 66
QY 283 -----HEDYHSELGDIDHDDHGVNESHRYSWINIFGDDSVLGSIATIH-ORDHLRK 336
DB 67 KENGAPEDETHNADGLNVTAGEDGVANINVTDSQIPLTGPNISITIGRAVVVHADPDLDLGR 126
QY 337 -----SAKIACCVIG 346
DB 127 GGHELKSTGNAGRVACGILG 148

RESULT 13
SODI_ORYSA STANDARD; PRT; 151 AA.
AC P28756;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 1 (EC 1.15.1.1).
GN SODCC1 OR SODCC.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RA MEDLINE=92322961; PubMed=1623183;
RA Sakamoto A., Ohsuga H., Tanaka K.;
RT Nucleotide sequences of two cDNA clones encoding different Cu/Zn-
RT superoxide dismutases expressed in developing rice seed (Oryza sativa
RT plant Mol. Biol. 19:323-327(1992)).
RN 120
RP SEQUENCE FROM N.A.
RC STRATGEV; Jaisong 67; TISSUE=Shoot;
RA Pao S.M., Huang G.B.;
RT Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
RN 131
RP SEQUENCE OF 1-90 FROM N.A.
RX MEDLINE=95241628; PubMed=7724677;
RX Sakamoto A., Okumura K., Kamakura H., Tanaka K.;
RT Molecular cloning of the gene (SODCC1) that encodes a cytosolic
RT copper/zinc-superoxide dismutase from rice (Oryza sativa L.).*;
RN 141
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;

```

```

CC -1- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: D00999; BA00799.1; -.
DR EMBL: U36320; AA03917.1; -.
DR EMBL: U19435; AAC1464.1; -.
DR PIR: S22508; S22508.
DR HSSP: P07505; 1SRD.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu.1.
DR PRINTS: PR00068; CUZNDISMASE.
DR PRODOM: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2.1.
DR Oxidoreductase; Copper; Zinc; Multigene family.
FW INIT_MET 0 0
FT METAL 44 46 COPPER (BY SIMILARITY).
FT METAL 46 46 COPPER (BY SIMILARITY).
FT METAL 61 61 COPPER AND ZINC (BY SIMILARITY).
FT METAL 69 69 ZINC (BY SIMILARITY).
FT METAL 78 78 ZINC (BY SIMILARITY).
FT METAL 81 81 ZINC (BY SIMILARITY).
FT METAL 118 118 COPPER (BY SIMILARITY).
FT DISULFID 55 144 COPPER (BY SIMILARITY).
FT SEQUENCE 151 AA; 15120 MW; CF5FD97ACD4E7998 CRC64;

Query Match
Best Local Similarity 32.4%; Pred. No. 1.3e-05; Length 151;
Matches 46; Conservative 19; Mismatches 51; Indels 26; Gaps 4;

QY 225 VHGTFKQVGYGDLGVSYHLEGFNVSDHKDHLVDQIYANGDITLSCDNLGAKYDP-- 282
DB 13 VKGTIFFTQEGDGPVAVTGSVGL-----KPLGHGFHVALDITNGCMSTGPHNPNAG 66
QY 283 -----HEDYHSELGDIDHDDHGVNESHRYSWINIFGDDSVLGSIATIH-ORDHLRK 336
DB 67 KENGAPEDETHNADGLNVTAGEDGVANINVTDSQIPLTGPNISITIGRAVVVHADPDLDLGR 126
QY 337 -----SAKIACCVIG 346
DB 127 GGHELKSTGNAGRVACGILG 148

RESULT 14
SODC_ZANAE STANDARD; PRT; 152 AA.
AC O65174;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] 1 (EC 1.15.1.1).
GN SODCC OR SOD3.
OS Zantedeschia aethiopica (White calla lily).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Zantedeschia.
OX NCBI_TaxID=69721;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;

```

RA Lino-Neto T., Tavares R.M., Palme K., Pais M.S.S.;  
RT "Expression of cytosolic Cu/Zn-superoxide dismutase during senescence  
and regression of Zanteleschschia spathe";  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Destroys radicals which are normally produced within the  
cells and which are toxic to biological systems.  
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -1- COFACTOR: Copper and zinc (By similarity).  
CC -1- SUBUNIT: HOMODIMER (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF054150; AAC08581.1; -  
CC HSSP: P07505; ISR0.  
CC Interpro: IPR001424; SOD\_CU\_ZN.  
CC Pfam: PF00080; sodcu; 1.  
CC PRINTS: PR00068; CUZNDISMTASE.  
CC Prodom: PD000469; SOD\_CU\_ZN; 1.  
CC PROSITE: PS00087; SOD\_CU\_ZN\_1; 1.  
CC PROSITE: PS00332; SOD\_CU\_ZN\_2; 1.  
CC Oxidoreductase; Copper; zinc.  
CC METAL 45 45 COPPER (BY SIMILARITY).  
CC METAL 47 47 COPPER (BY SIMILARITY).  
CC METAL 62 62 COPPER AND ZINC (BY SIMILARITY).  
CC METAL 70 70 ZINC (BY SIMILARITY).  
CC METAL 79 79 ZINC (BY SIMILARITY).  
CC METAL 82 82 ZINC (BY SIMILARITY).  
CC METAL 119 119 COPPER (BY SIMILARITY).  
CC METAL 145 145 BY SIMILARITY.  
CC DISULFID 56 145  
CC SEQUENCE 152 AA; 15152 MW; 96C535E190102160 CRC64;  
Query Match 5.8%; Score 163; DB 1; Length 152;  
Best Local Similarity 31.7%; Pred. No. 1.3e-05;  
Matches 45; Conservative 22; Mismatches 49; Indels 26; Gaps 4;  
QY 225 VHGTFIDFKVGYGDLVSYHLEGFVNSDDHKLHDVQIYANGDLTSGCDNLGAKYDP-- 282  
DB 14 VQGVTFEFOEGEGPTTIGSLGL-----KPLHGFHVALGDTNGCMSTGPHFNPAG 67  
QY 283 -----HEDYHSELGDLGDHDDHGVVNSHRYSWINIFGDDSVLGRSTAIH-QRDHLAK 336  
DB 68 KKHGAPEDDGNRRHAGDLGVTVGEDEGTNFTVDSQIPLTGLNSVGRVAVHADSDDLGK 127  
QY 337 -----SAKIACCVIG 346  
DB 128 GGHLSKTTGNAGRLACGVIG 149  
RESULT 15  
SODC\_CANAL STANDARD; PRT; 153 AA.  
AC 059924;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).  
GN SOD1.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99233600; PubMed=10216241;  
RA Hwang C.S., Rhee G., Kim S.T., Kim Y.R., Huh W.K., Baek Y.U.,

RA Kang S.O.;  
RT "Copper- and zinc-containing superoxide dismutase and its gene from  
Candida albicans";  
RL Biochim. Biophys. Acta 1427:245-255(1999).  
CC -1- FUNCTION: Destroys radicals which are normally produced within the  
cells and which are toxic to biological systems.  
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
CC -1- COFACTOR: Copper and zinc (By similarity).  
CC -1- SUBUNIT: HOMODIMER (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AF046872; AAC12872.1; -  
CC HSSP: P00445; IJCY.  
CC Interpro: IPR001424; SOD\_CU\_ZN.  
CC Pfam: PF00080; sodcu; 1.  
CC PRINTS: PR00068; CUZNDISMTASE.  
CC Prodom: PD000469; SOD\_CU\_ZN; 1.  
CC PROSITE: PS00087; SOD\_CU\_ZN\_1; 1.  
CC PROSITE: PS00332; SOD\_CU\_ZN\_2; FALSE\_NEG.  
CC Oxidoreductase; Copper; zinc.  
CC INIT\_MET 0  
CC METAL 46 46 COPPER (BY SIMILARITY).  
CC METAL 48 48 COPPER (BY SIMILARITY).  
CC METAL 63 63 COPPER AND ZINC (BY SIMILARITY).  
CC METAL 71 71 ZINC (BY SIMILARITY).  
CC METAL 80 80 ZINC (BY SIMILARITY).  
CC METAL 83 83 ZINC (BY SIMILARITY).  
CC METAL 120 120 COPPER (BY SIMILARITY).  
CC METAL 146 146 BY SIMILARITY.  
CC DISULFID 57 146  
CC SEQUENCE 153 AA; 16015 MW; 8BC67A2D17DCB0F CRC64;  
Query Match 5.8%; Score 163; DB 1; Length 153;  
Best Local Similarity 30.1%; Pred. No. 1.3e-05;  
Matches 44; Conservative 23; Mismatches 53; Indels 26; Gaps 4;  
QY 225 VHGTFIDFKVGYGDLVSYHLEGFVNSDDHKLHDVQIYANGDLTSGCDNLGAKYDPH 283  
DB 13 VQGVTFEFOEGEGPTTIGSLGL-----KPLHGFHVALGDTNGCMSTGPHFNPAG 67  
QY 284 -----EDYHSELGDLGDHDDHGVVNSHRYSWINIFGDDSVLGRSTAIH----- 329  
DB 68 KKHGAPEDDGNRRHAGDLGVTVGEDEGTNFTVDSQIPLTGLNSVGRVAVHADSDDLGK 127  
QY 330 -----QRDHLKSAKIACCVIG 349  
DB 128 KGFEDSKTTGNAGRLACGVIG 153

Search completed: November 29, 2002, 12:33:14  
JOB time : 29 secs



|  |   |   |              |  |  |
|--|---|---|--------------|--|--|
| Db   | 21  | DEGOCIDGQNKDDHHNDHNDHNDHDDDEDTMVAQCEMEERNPKMASLLNNHNGSIEL     | 80           |  |  |
| Oy   | 61  | SQKGGAYVLEHLHLGFTGTSDDHDDHNGLLHMLGDMSSAGCDSIGELYNAHPEKNADPG   | 120          |  |  |
| Db   | 81  | SQGGHGAUYLEHLHLGFTGTSDDHDDHNGLLHMLGDMSSAGCDSIGELYNAHPEKNADPG  | 140          |  |  |
| Oy   | 121   | DLGDLVDDRCGVNVNHHVATMLDIDGAPRTAEALIGHSMITLQGSHTPDADTPASIRACV  | 180          |  |  |
| Db   | 141   | DLGLDVLDDRCGVNVNHHVATMLDIDGAPRTAEALIGHSMITLQGSHTPDADTPASIRACV | 200          |  |  |
| Oy   | 181   | IGHGAKRPETAAALHNELEEDKTEYHACGDSVSENPORALNNHNGTIDPKQGVGDLE     | 240          |  |  |
| Db   | 201   | IGHGAKRPETAAALHNELEEDKTEYHACGDSVSENPORALNNHNGTIDPKQGVGDLE     | 260          |  |  |
| Oy   | 241   | VSTHLEGFNVSDKHDLHDVQIYANGDLTSGCDNLGAKYDPHEDEYHSELDDLDHDD      | 300          |  |  |
| Db   | 261   | VSTHLEGFNVSDKHDLHDVQIYANGDLTSGCDNLGAKYDPHEDEYHSELDDLDHDD      | 320          |  |  |
| Oy   | 301   | HGVVNESHRYSMINIFGDSVLSGRSIALIHORDHLHKSARKIACCVIGRQSHPEIVHAKC  | 360          |  |  |
| Db   | 321   | HGVVNESHRYSMINIFGDSVLSGRSIALIHORDHLHKSARKIACCVIGRQSHPEIVHAKC  | 380          |  |  |
| Oy   | 361   | VVPRNTESTGLNNHNSGSIITEQTPGGSTHMTADLKGFNVSBDLSHNNHNGQLHEMGDMS  | 420          |  |  |
| Db   | 381   | VVPRNTESTGLNNHNSGSIITEQTPGGSTHMTADLKGFNVSBDLSHNNHNGQLHEMGDMS  | 440          |  |  |
| Oy   | 421   | HGHSHSGRMVHGHDDAHNPKRPGLDGVINDSGIYHSTRTEPDHNLVNDLNAKSLVIMOG   | 480          |  |  |
| Db   | 441   | HGHSHSGRMVHGHDDAHNPKRPGLDGVINDSGIYHSTRTEPDHNLVNDLNAKSLVIMOG   | 500          |  |  |
| Oy   | 481   | GHEVESERVAACCVIGRA  | 497          |  |  |
| Db   | 501   | GHEVESERVAACCVIGRA  | 517          |  |  |
| RESULT 2   |   |   |              |  |  |
| Q9NE57   | Q9NE57  | PRELIMINARY:  | PRT: 735 AA. |  |  |
| AC   | Q9NE57  |   |              |  |  |
| DT   | 01-OCT-2000 (TREMBLrel. 15.                                 | Created)  |              |  |  |
| DT   | 01-OCT-2000 (TREMBLrel. 15.                                 | Last sequence update)   |              |  |  |
| DT   | 01-DEC-2001 (TREMBLrel. 19.                                 | Last annotation update)                                       |              |  |  |
| DE   | Y39B6B_99 Protein (Y39B6A_pp                                | protein).   |              |  |  |
| GN   | Y39B6B.GG OR Y39B6A.pp.                                     |   |              |  |  |
| OS   | Caenorhabditis elegans.                                     |   |              |  |  |
| OC   | Eukaryota; Metazoa; Nematoda;                               | Rhabditida; Rhabditidae;                                      |              |  |  |
| OC   | Rhabditidae; Pelodierinae;                                  | Caenorhabditis.   |              |  |  |
| OX   | NCBI_TaxID=6239;  |   |              |  |  |
| RN   | RN  |   |              |  |  |
| RA   | SEQUENCE FROM N.A.  |   |              |  |  |
| RA   | Sulston J.E.;   |   |              |  |  |
| RL   | Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.     |   |              |  |  |
| RN   | [2]   |   |              |  |  |
| RP   | SEQUENCE FROM N.A.  |   |              |  |  |
| RX   | MEDLINE=99069613; PubMed=9851916;                           |   |              |  |  |
| RA   | none;   |   |              |  |  |
| RT   | "Genome sequence of the nematode C. elegans: A platform for | investigating biology."                                       |              |  |  |
| RL   | Science 282:2012-2018(1998).                                |   |              |  |  |
| RN   | [3]   |   |              |  |  |
| RP   | SEQUENCE FROM N.A.  |   |              |  |  |
| RA   | Sulston J.;   |   |              |  |  |
| RL   | Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.     |   |              |  |  |
| DR   | EMBL; AL132986; CAC6098.1; -                                |   |              |  |  |
| DR   | EMBL; AL132946; CAC51077.1; -                               |   |              |  |  |
| DR   | HSSP; P13231; IHCE.   |   |              |  |  |
| DR   | InterPro; IPRO02395; Kintogen.                              |   |              |  |  |
| DR   | PRINTS; PR00334; KINTOGEN.                                  |   |              |  |  |
| SO   | SEQUENCE 735 AA; 80235 MW; 6E7B831DBE41159 CRC64;           |   |              |  |  |
| Query Match  |   |   |              |  |  |
| Best Local Similarity 7.3%; Score 204.5; DB 5; Length 735; |   |   |              |  |  |
| 21.0%; Pred. No. 6.3e-08;                                  |   |   |              |  |  |

|    |   |   |   |                     |             |      |        |      |      |    |
|----|---|---|---|---------------------|-------------|------|--------|------|------|----|
|    | Matches   | 102;  | Conservative                                | 29;                 | Mismatches  | 178; | Indels | 177; | Gaps | 23 |
| OY | 1   | DREOSNDQKDDH--  | -IDDNDHDNDODDDDEDTMTVACSEMRPNMNASLNNHNSI    | 58                  |             |      |        |      |      |    |
| Dd | 386   | EGBNENNAANDENGVNHNNGNNGSTNN-----                            | SPANGBNHGSTNNNGEN                           | 432                 |             |      |        |      |      |    |
| OY | 59  | ELSQSGKAGVLTETLVGFTSDESDNH-----                             | NHGILMLMDMSAGDSIGELYN                       | 110                 |             |      |        |      |      |    |
| Dd | 433   | NNRANH--  | NHESGNHNSPRANGNNGENNARAHNGN-----            | GENNN               | 476         |      |        |      |      |    |
| OY | 111   | ANREKHADRGDLVDDDRGVNEVNNHYTMLDIDGTARTETALYGHSNTLIQSHTRAD    | 170   |                     |             |      |        |      |      |    |
| Dd | 477   | A-RANNNGENG-----  | TNN-----                                    | HSH---H             | 498         |      |        |      |      |    |
| OY | 171   | TRASIAACSVIGHKARETAALINHELEDKTENA-----                      | NCDVSNTYORALINN                             | 224                 |             |      |        |      |      |    |
| Dd | 499   | SRPH-----   | NHNGENNNARRAHNGNNGSTNNNGENNARRAHNGNNGENSTNG | 550                 |             |      |        |      |      |    |
| OY | 225   | VNOTIDFKVGUDLEVSYHLEGFNVSDIKRLDYQVANGDLTSGCDNLGAKYDPHE      | 284   |                     |             |      |        |      |      |    |
| Dd | 551   | NHGS-----   | NHSPRANHNGENNARRAHNGNNGN-----               | SGSVGNNGNE          | 591         |      |        |      |      |    |
| OY | 285   | DY-NSELGLDGDINDDRGVNESHRSWINEFGDSDVLGRSLAINORDHLKSATACS     | 343   |                     |             |      |        |      |      |    |
| Dd | 592   | SHNGNRRANNGHNGEIVNGNNGN-----                                | GAGYGANNGN-----                             | 627                 |             |      |        |      |      |    |
| OY | 344   | VIGRGGSHREIYNRAKCVVRPTTESTCLNNHSSGSIFFEEPCGSTNMTADLKGFVNSED | 403   |                     |             |      |        |      |      |    |
| Dd | 628   | ----GANHHNARH-----  | NEHENNEGDNHN--                              | SGSHVNGHNGSTNN----- | S           | 663  |        |      |      |    |
| OY | 404   | LSHNRGYSLDHNMCSMGSCSLGMRYNGRDADRPKRPDLDVDLIDDSHGAYSTRPD     | 463   |                     |             |      |        |      |      |    |
| Dd | 664   | LANNH-----  | HCGSTNNGANN-SPRANHNGNCAH-----               | NEGANNGANNH         | 705         |      |        |      |      |    |
| OY | 464   | HLAVED  | 469   |                     |             |      |        |      |      |    |
| Dd | 706   | HDKEN   | 711   |                     |             |      |        |      |      |    |
|    | RESULT 3  |   |   |                     |             |      |        |      |      |    |
|    | 09QZV4  | PRELIMINARY:  | PRT:  | 726 AA.             |             |      |        |      |      |    |
| ID | 09QZV4  |   |   |                     |             |      |        |      |      |    |
| AC | 09QZV4;   |   |   |                     |             |      |        |      |      |    |
| DT | 01-MAY-2000 (TREMBREL. 13,  | Created)  |   |                     |             |      |        |      |      |    |
| DT | 01-MAY-2000 (TREMBREL. 13,  | Last sequence update)                                       |   |                     |             |      |        |      |      |    |
| DT | 01-JUN-2001 (TREMBREL. 17,  | Last annotation update)                                     |   |                     |             |      |        |      |      |    |
| DE | Histidine-rich Ca2+ binding protein.                              |   |   |                     |             |      |        |      |      |    |
| GN | HRC.  |   |   |                     |             |      |        |      |      |    |
| OS | Mus musculus (Mouse).   |   |   |                     |             |      |        |      |      |    |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |   |   |                     |             |      |        |      |      |    |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.         |   |   |                     |             |      |        |      |      |    |
| OX | NCBI_TaxID=10090;   |   |   |                     |             |      |        |      |      |    |
| RN | [1]   |   |   |                     |             |      |        |      |      |    |
| RP | SEQUENCE FROM N.A.  |   |   |                     |             |      |        |      |      |    |
| RC | TISSUE=SKELETAL MUSCLE;   |   |   |                     |             |      |        |      |      |    |
| RX | MEDLINE=93389148; PubMed=10462052;                                |   |   |                     |             |      |        |      |      |    |
| RA | Ridgeway A.G., Petropoulos H., Siu A., Ball J.K., Skerjanc I.S.;  |   |   |                     |             |      |        |      |      |    |
| RT | "Cloning, tissue distribution, subcellular localization and       |   |   |                     |             |      |        |      |      |    |
| RT | overexpression of murine histidine-rich Ca2+ binding protein.";   |   |   |                     |             |      |        |      |      |    |
| RL | Fors Lett. 456:399-402(1999).                                     |   |   |                     |             |      |        |      |      |    |
| DR | EMBL: AF158597; AAD5250.1; -                                      |   |   |                     |             |      |        |      |      |    |
| DR | MGI: MGI:166226; RfC.   |   |   |                     |             |      |        |      |      |    |
| DR | InterPro: IPR000561; EGF-like.                                    |   |   |                     |             |      |        |      |      |    |
| DR | InterPro: IPR002049; Laminin_EGF.                                 |   |   |                     |             |      |        |      |      |    |
| DR | PROSITE: PS00022; EGF_1; UNKNOWN_1.                               |   |   |                     |             |      |        |      |      |    |
| DR | PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.                    |   |   |                     |             |      |        |      |      |    |
| SO | SEQUENCE 726 AA; 83623 MW; 6B13ZF250C5904B30 CRC64;               |   |   |                     |             |      |        |      |      |    |
|    | Query Match   | 6.4%;   | Score 178.5;                                | DB 11;              | Length 726; |      |        |      |      |    |
|    | Best Local Similarity   | 21.5%;  | Pred. NO. 7.6e-06;                          |                     |             |      |        |      |      |    |
|    | Matches 118;  | Conservative 62;  | Mismatches 189;                             | Indels 181;         | Gaps 29;    |      |        |      |      |    |
| OY | 2   | GEOSNDQAKNDHNDH--   | -NDHNDHDDDDDEDTMTVACSEMRPNMNASLNNHNY----    | H                   | 55          |      |        |      |      |    |

```
Db 117 GHENEDNEDLGSAENHLPRORSHSHEDDGI-----VSSSEYHHRVPRNAH 163
QY 56 GSTELISOKHGCVUYLELILVGFNTSEDDHNNHGLHLMLGDMAGCDSIGELYNAHPEK 115
Db 164 GHGEEDDDDDGEEEEE-----RVDMEDSDDNEHQVHGH-----QSHSKREDELHNAHSHR 214
QY 116 HADPGGLGYLDORGVVNE-----VHHYAMLDDIGTAPRTFALIGHSMTLLGSGTADPT 171
Db 215 HGGHSD-----DDDDGVSTENGQAHNYQHNEEDDSDSDSHTRVQGREDENDDEG 270
QY 172 PASRIACCVIGKARPTAALH-----ELEEDKTEYAHCDVSNTHQ 217
Db 271 DS-----GEYRHTQDHQNEEDDDDDDDDEKEDSTEH-----RHQ 311
QY 218 PRALH-----HHV--HGTIDKQVGTGLEVSYHLEGFVSDKHLDHVOI 263
Db 312 TQGNKKEDEDESDDDHVSRRHG-----RQGEDEEDDDDDG--DDSTENHQAHR 363
QY 264 YANGDLTSGCDNLGAKYRPHEDYSELGLDIDHDDHGVVNESHRYSWNIIFGDSVLG 323
Db 364 HRDHEKDED-----DSEEDYH-----HVPSHG--RQSHQ--NEEDEAV-----STEMH 401
QY 324 RSTALHQRDLHKSAKIAACVIGRQSHPEIYHRAKCV--RP-----NTSTGLHHV 375
Db 402 -----STEMHQAHRNAHNLGR--ESEEEVAVKYSHHVAHSRPOGNADREDSLEEHM 454
QY 376 SGIPTFQTPGGSTMTADLKGFVSEDLSHNRHGVQLHEMGDMHSGHSLGRMYHGHD 435
Db 455 N-----EVRGHHNHRAHSGDEEDISTEFGHKAPSHRLQODE-----RARGHRE 499
QY 436 -----AHDPKRP-----GDLDVYIDSHGIVHSTFEDHUNVEDLNARSLYIM 478
Db 500 RVQGEIHAQRLQPTGSPRSRESKRED-----HSSQSGD-----EDPEQR----- 538
QY 479 OGGHEVESER 488
Db 539 QAHSEEEEX 548

RESULT 4
QY 09WE4 PRELIMINARY: PRT: 738 AA.
AC 09WE4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Histidine-rich calcium-binding protein.
GN HRC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HEART;
RA Shi S.D., Brunner S.R.;
RT "Molecular cloning of histidine-rich calcium-binding protein from
RT mouse hearts."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF132218; AAD42061.1;
DR MGD; MGI:96226; HRC.
DR InterPro: IPR000561; EGF-like
DR InterPro: IPR002049; Laminin_EGF.
DR PROSITE: PS01242; EGF_1; UNKNOWN_1.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
SQ SEQUENCE 738 AA: 85004 MW: D1F6ZDE2D533ECA2 CRC64;

Query Match 6.3%; Score 177; DB 11; Length 738;
Best Local Similarity 21.8%; Pred. No. 1e-05;
Matches 105; Conservative 56; Mismatches 180; Indels 140; Gaps 24;
QY 18 HNDHHDHDDHDDDETMHY-----AQCEMERPRHMASSLHHNV-----HGSIELSQHGCA 67
```

```
Db 116 HNGHEDNEDLDGSAENHLPRORSHSHEDDGIYSSEYHGVPRNAHNGHGEEDDDG 175
QY 68 VYLELILVGFNTSEDDHNNHGLHLMLGDMAGCDSIGELYNAHPEKHAHDPGDLVD 127
Db 176 EEEE-----RVDMEDSDDNEHQVHGH-----QSHSKREDELHNAHSHRQHSO-----DD 222
QY 128 DRGVVNE-----VHHYAMLDDIGTAPRTFALIGHSMTLLGSGTADPTPASRIACCVIGH 183
Db 223 DDDGVSTENGQAHNYQHNEEDDSDSDSHTRVQGREDENDDEG 272
QY 184 GARRETAALH-----ELEEDKTEYAHCDVSNTHQKALH----- 222
Db 273 GEYRHTQDHQNEEDDDDDDDDEKEDSTEH-----RHQQGNKKEDEDE 323
QY 223 -----HHV--HGTIDKQVGTGLEVSYHLEGFVSDKHLDHVOIYANGDLTSGDN 275
Db 324 SDEDDHVSRRHG-----RQGEDEEDDDDDG--DDSTENHQAHRHREKDED- 374
QY 276 LQAKYRPHEDYSELGLDIDHDDHGVVNESHRYSWNIIFGDSVLGRSIALHQRDLH 335
Db 375 -----DSEEDYH-----HVPSHG--RQSHQ--NEEDEAV-----STEMH 407
QY 336 KSAKIAACVIGRQSHPEIYHRAKCV--RP-----NTSTGLHHVSGSITFQTPGG 387
Db 408 QSPRIARNDLGR--ESEEEVAVKYSHHVAHSRPOGNADREDSLEEHM-----EVRGH 460
QY 388 STMTADLKGFVSEDLSHNRHGVQLHEMGDMHSGHSLGRMYHGHD-----AHDPKRP 441
Db 461 HNHRAHSGDEEDISTEFGHKAPSHRLQODE-----RARGHREPVQGEIHAQRLQ 511
QY 442 P 442
Db 512 P 512

RESULT 5
QY 064466 PRELIMINARY: PRT: 251 AA.
AC 064466;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Superoxide dismutase (Cu-Zn) (EC 1.15.1.1) (Superoxide dismutase 3,
DE extracellular).
GN SOD3 OR SOD 3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=LIVER;
RX MEDLINE-95327627; PubMed-7603981;
RA Carlsson L.M., Jonsson J., Edlund T., Marklund S.M.;
RT "Mice lacking extracellular superoxide dismutase are more sensitive to
RT hyperoxia."
RL Proc. Natl. Acad. Sci. U.S.A. 92:6264-6268 (1995).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=LIVER;
RX MEDLINE-9624263; PubMed-8643556;
RA Carlsson L.M., Marklund S.M., Edlund T.;
RT "The rat extracellular superoxide dismutase dimer is converted to a
RT tetramer by the exchange of a single amino acid."
RL Proc. Natl. Acad. Sci. U.S.A. 93:5219-5222 (1996).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Neplionev I.V., Folz R.J.;
RT "Mouse Extracellular Superoxide Dismutase (SOD3) gene."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
[4]
```

RP SEQUENCE FROM N.A.  
RA Strassberg R.;  
CC Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
DR EMBL; X84940; CAA59335.1; -.  
DR EMBL; AF039602; AAC62204.1; -.  
DR EMBL; BC010975; AAH10975.1; -.  
DR HSSP; P15107; 1XSO.  
DR MGD; MGI:103181; Sod3.  
DR InterPro; IPR001424; SOD\_CU\_ZN.  
DR Pfam; PF00080; sodcu; 1.  
DR PRINTS; PR00068; CUZNDISMUTASE.  
DR PRODOM; PD000469; SOD\_CU\_ZN.1.  
DR PROSITE; PS00087; SOD\_CU\_ZN.1; 1.  
DR PROSITE; PS00332; SOD\_CU\_ZN.2; 1.  
KW Copper; Oxidoreductase; Zinc.  
SQ SEQUENCE 251 AA; 27409 MW; 27D26EAL610ZCAE CRC64;  
  
Query Match 6.2%; Score 174.5; DB 11; Length 251;  
Best Local Similarity 31.1%; Pred. No. 3.8e-06;  
Matches 56; Conservative 22; Mismatches 57; Indels 45; Gaps 11;  
  
QY 199 EEDKTEHYAHCDYRSNTNPKAL-----HHHVGITDFKVGTD-LEVSYHLGSEFNVSD 253  
DB 67 EVDAAEMHAIARV-----QPSATLPDPDQPTGLVLFKRLPGSRLEAYFSLGEPAEQN 121  
QY 254 HKD---HLHDVQIYANGDLTSGCDNLGAKYDPHEDYHSELGLDIDDDHG--VYNES- 307  
DB 122 ASNNAIVHFE-----GDSQGDSTGPRYINMEVPHDQ-----HKGDFGNVYVANGQ 169  
QY 308 ---HRVS-WINIFGDSVLRSTAIHO--RDHLKSA-----KIACCVIGRGS 350  
DB 170 LMRHVRGLTSLAGPHSILGRSVYVAGEDLGKGNQASLQNGNAGRRLACCVYGTSS 229  
  
RESULT 6  
Q9ZNO4 PRELIMINARY; PRT; 152 AA.  
AC Q9ZNO4;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Superoxide dismutase (Ec 1.15.1.1) (Cu-Zn).  
GN SODCC OR SOD-1.  
OS Cicer arietinum (chickpea) (Garbanzo).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Oxysids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.  
OX NCBI\_TaxID=827;  
RA SEQUENCE FROM N.A.  
RA STRAIN-CV. ILC3279; TISSUE-LEAF;  
RA Hanselle T., Barz W.;  
RT "Isolation of cDNA Encoding Superoxide Dismutase from Infected  
RT Chickpea Plants."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. CASTELLANA; TISSUE-ETIOLATED EPICOTYLS;  
RA Dopico B., Hernandez-Nistal J., Labrador E.;  
RT "A cDNA encoding a superoxide dismutase is expressed in chickpea  
RT epicotyls."  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
DR EMBL; A0102739; CA10160.1; -.  
DR EMBL; A0102691; CA10132.1; -.  
DR HSSP; P07505; ISR0.

DR InterPro; IPR001424; SOD\_CU\_ZN.  
DR Pfam; PF00080; sodcu; 1.  
DR PRINTS; PR00068; CUZNDISMUTASE.  
DR PRODOM; PD000469; SOD\_CU\_ZN.1.  
DR PROSITE; PS00087; SOD\_CU\_ZN.1; 1.  
DR PROSITE; PS00332; SOD\_CU\_ZN.2; 1.  
KW Copper; Oxidoreductase; Zinc.  
SQ SEQUENCE 152 AA; 15222 MW; 4ED0AF0A54A8A454 CRC64;  
  
Query Match 6.2%; Score 174; DB 10; Length 152;  
Best Local Similarity 33.1%; Pred. No. 2.1e-06;  
Matches 47; Conservative 23; Mismatches 46; Indels 26; Gaps 4;  
  
QY 225 VHGTFKQYGVGDLEVSYNLGEFNVSDHKRDHLHDVQIYANGDLTSGCDNLGAKYDPH- 283  
DB 14 VSGTINSGQGDGPTVTGNTLGL-----KPLGHPHIALGDTMGCTSTGPHNPNMG 67  
QY 284 -----EDYHSELGLDIDDDHGUVNESHRYSWINIFGDDSVLRSTAIH--QRPDLHK 336  
DB 68 KENGSPEDPIRHHAGDLGNINVGDDGTVSFSITDNLPLTGPNSTIGRAVYVHADPDGLCK 127  
QY 337 -----SAKIACCVIG 346  
DB 128 GGHBLKTTGNAGRYACGIG 149  
  
RESULT 7  
Q90831 PRELIMINARY; PRT; 1840 AA.  
AC Q90831;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Translatin.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-95300324; PubMed-7781165;  
RA Kelly M.M., Phanthourath C., Breas D., McCabe C.F., Cole G.J.;  
RT "Molecular characterization of RAP-300: a high molecular weight  
RT embryonic polypeptide containing an amino acid repeat comprised of  
RT multiple leucine-zipper motifs."  
RL Brain Res. Dev. Brain Res. 85:31-47(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Cole G.J.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X60877; CAA56845.1; -.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; Filament; 1.  
DR PROSITE; PS00226; IF; UNKNOWN\_1.  
SQ SEQUENCE 1840 AA; 202279 MW; 8ED355F3F474831 CRC64;  
  
Query Match 6.2%; Score 173; DB 13; Length 1840;  
Best Local Similarity 22.7%; Pred. No. 7.3e-05;  
Matches 111; Conservative 60; Mismatches 175; Indels 142; Gaps 32;  
  
QY 11 KDHNND--DHNDNDHDDDDDETMNVASQSEMERPNHMASSLNHNHVGSTL-----SQKG 64  
DB 825 REEHRLDQEGNGDLQVENEHDEWEGDIOQE-----HGDTQEEVGTQEE 869  
QY 65 HGAUVLELNHUGNTESEHDD--NNHGLNLNMGASCSIGELYNAHREKNAPRGL 122  
DB 870 HG-----DLQEGSGDLQENHGDQVQENHGDQVQENHGDQVQENHGDQVQENHGD 922  
QY 123 -----GDL-----VDDRGVNVENHNAW-----LDLQARPTALNHSMTLQSGHTDA 169  
DB 923 QEHNGDLQVENEHDEWEGDIOQENHDEWEGDIOQENHDEWEGDIOQENHDEWEGDIO 978

QY 170 DPAASITACCVGHGKAPRETAALAHNELEE--DKREYAHNCVSRNTQPALNHNH 226  
Db 979 QVSGDGLQ--BEHGDQGE-----HEDLDQEHGDTQEHGDLQVHNGDQEE--HRDY- 1027  
QY 227 GTIDFQVQVGYGLEYSY--HLEGPNVSDKHKLHVOUYATANDLTSQDNIGATYD-- 281  
Db 1028 -----QEHGDLQVHNGDLQVHNGDLQEHNGDQEE-----HDDLQEHNGDQEHNGDL 1076  
QY 282 --PHEHYSEILD-----GDHDDHGVVNSHRYSMINIFGDSVYLSRAIHNQDHL 335  
Db 1077 QVHEHDLQVHNGDLQVHNGDL--QEHNGDTQEEH--DLQVHNGDQEEHRLD--QEHNGD 1131  
QY 336 KSAKICCVYIGGQSHPELVYRPAKVVRPNTESGLHHNVSGSTTEPQPGSGTMTADL 395  
Db 1132 -----LQEHGDLQVH-----GDLQVH-----GDL 1153  
QY 396 KGFNVESDLSHNR-----HGVOLEHMGMSHGCHSLGRMYGHNDADHPKRPGLD---G 446  
Db 1154 QD-EYGDGTQEEHRLDQVHNGDQEEHRLDQEG-H--GDLQEGHDLME--EHGDLKEENG 1207  
QY 447 DVIDSHG 454  
Db 1208 D-LQEGHG 1214

RESULT 8  
088592  
ID ID 088592 PRELIMINARY; PRT: 251 AA.  
AC 088592;  
DT 01-NOV-1998 (TREMblrel. 08, Created)  
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Extracellular superoxide dismutase.  
GN SOD3  
OS Mus musculus (Mouse)  
OC Pakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutaria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=APQ(E)-DEFICIENT MOUSE; TISSUE=ATHEROSCLEROTIC VESSEL;  
RZ MEDLINE=98526350; PubMed=9593766;  
RA Fukui T., Galls Z.S., Meng X.P., Parthasarathy S., Harrison D.G.;  
RT Vascular expression of extracellular superoxide dismutase in  
RT atherosclerosis.  
RT J. Clin. Invest. 101:2101-2111(1998).  
RL - FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).  
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
CC EMBL: AF076286; JAC26522.1; -  
DR HSSP: P00445; ITCV.  
DR MGD: MGI:103181; Sod3.  
DR InterPro: IPR001424; SOD\_CU\_ZN.  
DR Pfam: PF00080; sodcu.1.  
DR PRINTS: PR00068; CUZNDISMASE.  
DR PRODOM: PD000469; SOD\_CU\_ZN.1.  
DR PROSITE: PS00087; SOD\_CU\_ZN.1.1.  
DR PROSITE: PS00332; SOD\_CU\_ZN.2.1.  
KW Copper; Oxidoreductase; Zinc.  
KW SEQUENCE 251 AA, 27393 MW, 27CA16C197AAMAE CRC64;

Query Match 6.1%: Score 171.5; DB 11; Length 251;  
Best local similarity 30.6%: Pred. No. 6.5e-06;  
Matches 55; Conservative 23; Mismatches 57; Indels 45; Gaps 11.

QY 199 EEKTEHYAHCDYRSTWHPKAL-----HHHVGITDFKQVGYGD-LEVSYHLEGPNVSD 253  
Db 67 EVDAAEHNAICR-----QPSATLPDPDQPTGLVLEKQDAPGSRLLATVYLSGFPAEON 121  
QY 254 HKD---HLHDVQIYANGDLTSCDNLGAKYDPHEDYHSELGDLGDHDDHG--VYNE- 307  
Db 122 ASRNAIHVHEF-----GDLISGCDSTGYNHDMVEYHQ-----HPGFGNFWANQO 169

| ID                       | Q0908H0  | PRELIMINARY: | PRT: | 163 AA. |
|--------------------------|--|--------------|------|---------|
| AC                       | Q0908H0:   |              |      |         |
| DT                       | 01-MAY-2000 (Tremblrel. 13, Created)                                   |              |      |         |
| DT                       | 01-MAY-2000 (Tremblrel. 13, Last sequence update)                      |              |      |         |
| DT                       | 01-MAY-2002 (Tremblrel. 20, Last annotation update)                    |              |      |         |
| DE                       | M31R.  |              |      |         |
| GN                       | M31R.  |              |      |         |
| OS                       | Myxoma virus (strain Lausanne).  |              |      |         |
| OC                       | Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;    |              |      |         |
| OC                       | Leporipoxvirus.  |              |      |         |
| OX                       | NCBI_TaxID=31530;  |              |      |         |
| RN                       | 11   |              |      |         |
| RP                       | SEQUENCE FROM N.A.   |              |      |         |
| RC                       | STRAIN-LAUSANNE:   |              |      |         |
| RA                       | Cameron C., Hota-Mitchell S., Chen L., Barrett J., Cao J.-X.,          |              |      |         |
| RA                       | Macaulay C., Miller D., Evans D., McPadden G.;                         |              |      |         |
| RA                       | Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.                |              |      |         |
| CC                       | -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE |              |      |         |
| CC                       | CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).             |              |      |         |
| CC                       | -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) = O(2) + H(2)O(2).       |              |      |         |
| CC                       | -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.      |              |      |         |
| DR                       | EMBL: AF170726; AAF15019.1; -.   |              |      |         |
| DR                       | HSSP: P00441; 4SOD.  |              |      |         |
| DR                       | Interpro: IPR001424; SOD_CU_ZN.  |              |      |         |
| DR                       | Pfam: PF00080; sodcu.1.  |              |      |         |
| DR                       | PRINTS: PR00068; CUZNDISMASE.  |              |      |         |
| DR                       | Prodom: PD000469; SOD_CU_ZN.1.   |              |      |         |
| KW                       | Copper; Oxidoreductase; Zinc.  |              |      |         |
| SD                       | SEQUENCE 163 AA: 17786 MW; 0C48CC745DFAT7A CRC64;                      |              |      |         |
| Query Match              | 6.1%; Score 170.5; DB 12: Length 163;                                  |              |      |         |
| Best Local Similarity    | 28.9%; Pred. No. 4.4e-06;  |              |      |         |
| Matches 58; Conservative | 28; Mismatches 38; Indels 77; Gaps 12;                                 |              |      |         |
| OY                       | 199 EEDKTEHYAHCDVRSNTHPKALHHNHVHTDIDFKOVGYGDLVSYNL-----EGFNVSDD 253    |              |      |         |
| DB                       | 11 EFNKVKRAKACMLRGN-----DYKVLNHFEDL-ONDIVVVFVGVLTLTGSGF----- 55        |              |      |         |
| OY                       | 254 HKDLIHVOJUYANGDLTSGCDN-LGAKUYDPHE-----DVHSELGLADGIDHDDGGVYN 305    |              |      |         |
| DB                       | 56 -----HGINYHEFGD---ACSNELGCHYNEPKRSHGAPSDVNRHVRDGLGNHIAHSGKV-- 105   |              |      |         |
| OY                       | 306 ESHRSWYNLF-----GDDSVLGSRSLAHORDLHKSAKIACVYGRGQSPHELVHRA 358        |              |      |         |
| DB                       | 106 -----SVYHIIDKILTLTGDRSLIGSLAHHRD-----DLGRG----- 140                |              |      |         |
| OY                       | 359 KCVYRPMTSTGLHHVSGSI 379  |              |      |         |
| DB                       | 141 -----FVWESS--IHGNSGV 154   |              |      |         |
| RESULT 10                |  |              |      |         |
| ID                       | Q25878   | PRELIMINARY: | PRT: | 430 AA. |
| AC                       | Q25878:  |              |      |         |
| DT                       | 01-NOV-1996 (Tremblrel. 01, Created)                                   |              |      |         |
| DT                       | 01-NOV-1996 (Tremblrel. 01, Last sequence update)                      |              |      |         |





```
RL Genetics 144:317-328(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-90136495; PubMed-2482436;
RA Cannon R.E., Scandalios J.G.;
RT "Two cDNAs encode two nearly identical Cu/Zn superoxide dismutase
   proteins in maize."
RL Mol. Gen. Genet. 219:1-8(1989).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
   CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL: U34727; AAB49913.1; -.
DR EMBL: X17564; CAB57993.1; -.
DR HSSP: P07505; 1SRD.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu; 1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00087; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
KM Copper: Oxidoreductase; Zinc.
SQ SEQUENCE 152 AA; 15089 MW; 691DFF62F88A7623 CRC64;

Query Match 5.8%; Score 163; DB 10; Length 152;
Best Local Similarity 31.7%; Pred. No. 1.6e-05;
Matches 45; Conservative 21; Mismatches 50; Indels 26; Gaps 4;

QY 225 VHGTFDKQVGGDLEVSYLEGFVNSDDHKDHLHDVOIYANGDLTSGCDNLGAKYDP-- 282
   |||||
DB 14 VKDTFFTDGDDPTAVNTVSAGL-----KPGHLGFVHVALGTTNGCKSTGHDYNPAS 67
   |||||

QY 283 -----HEDYHSELGDIHDDHGVNESHRYSMINIFGDDSVLGRSIAIR-ORDHLK 336
   |||||
DB 68 KERGAPDEENRAGDIGNVTAGADANINVTDSOIPLTGPNSTIRGAVVHADPDLDLK 127
   |||||

QY 337 -----SAKIACCVIG 346
   :|||
DB 128 GGHLSKSTGNAGRVACGILG 149

RESULT 13
08MNN6 PRELIMINARY; PRT; 153 AA.
ID 08MNN6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cu/Zn superoxide dismutase.
OS SOD1.
OC Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxId=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Green S.L., Tolwani R.J., Varma S., Quignon P., Galibert F.,
RA Cork L.C.;
RT "The canine Cu/Zn superoxide dismutase gene (SOD1).";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF346417; AAL61608.1; -.
DR InterPro: IPR001424; SOD_CU_ZN.
DR Pfam: PF00080; sodcu; 1.
DR PRINTS: PR00068; CUZNDISMUTASE.
DR ProDom: PD000469; SOD_CU_ZN.1.
DR PROSITE: PS00332; SOD_CU_ZN.2; 1.
SQ SEQUENCE 153 AA; 15912 MW; 0D7900E59C57E6B0 CRC64;

Query Match 5.8%; Score 163; DB 6; Length 153;
Best Local Similarity 31.7%; Pred. No. 1.6e-05;
Matches 46; Conservative 22; Mismatches 51; Indels 26; Gaps 5;

QY 225 VHGTFDKQVGGDLEVSYLEGFVNSDDHKDHLHDVOIYANGDLTSGCDNLGAKYDP-- 282
```

```
DB 15 VECTHFFVKGSGPVVSGTITGL-TEGEHGFVHGF-----DXTGCTSGAPHPNPLS 68
   |||||
QY 283 -----HEDYHSELGDIHDDHGVNESHRYSMINIFGDDSVLGRSIAIRH-ODHLK 336
   |||||
DB 69 KHGKGRDERRVGDGNVTAGKDCVAVSTEDSLTNSDYSIIIRTVHVEKKRDLGK 128
   |||||

QY 337 -----SAKIACCVIGRQ 349
   :|||
DB 129 GDNESTGTGNAGSLACGVIGIQ 153

RESULT 14
Q9V523 PRELIMINARY; PRT; 179 AA.
ID Q9V523;
AC Q9V523;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CG9027 protein.
GN CG9027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephyroptera; Neoptera; Emdopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephyroptidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Bailew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostali D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Daitali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshireli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirding A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
   CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
DR EMBL: A8003826; AAF58647.2; .
```

DR HSP, P07505; 1SRD.  
 DR FLYBase; FBgn0033631; CG9027.  
 DR InterPro; IPR001424; SOD\_CU\_ZN.  
 DR Pfam; PF00080; SODcu; 1.  
 DR PRINTS; PR00068; CUZNDISMTASE.  
 DR PRODOM; PD000469; SOD\_CU\_ZN; 1.  
 DR PROSITE; PS00087; SOD\_CU\_ZN; 1.  
 DR PROSITE; PS00332; SOD\_CU\_ZN; 2; 1.  
 DR Copper; Oxidoreductase; Zinc.  
 DR SEQUENCE 179 AA; 18950 MW; 8C45D94E22D81B9F CRC64;  
 Query Match 5.8%; Score 162.5; DB 5; Length 179;  
 Best Local Similarity 27.6%; Pred. No. 2.2e-05;  
 Matches 53; Conservative 31; Mismatches 61; Indels 47; Gaps 9;  
 DB 333 HLKRSKACVIGRGSGHP--EIVHRAKCVVRPMTSTGILHNHVSGLTFEOTP--GGST 389  
 4 YLVSLALCATICSAQTNPICAIAYLIGPVQSDNTQ-----VKGNVFTQNDCCGVN 56  
 QY 390 HMTDLGFWNSDELSHRHGCVGLHMGSMHSGHSLGRMY-----HGDDAHDPKRP 443  
 DB 57 HVAVQLG-----LKEKGKGFHHEKGLTNGCISMGAHNPKVDIGCPD-HEVHVHG 109  
 QY 444 DLGVINDSHGTVHSTFPHLWVE--DLNRSYVIMOGCHEVSE----- 487  
 DB 110 DLGNLEANSIGIIVYITTDVYITLTGKLGITIGRVVY-----HELEDGLGHNHDSKKT 165  
 QY 488 ----RVACVYG 495  
 DB 166 NAGGRIACGVIG 177  
 RESULT 15  
 Q917F0 PRELIMINARY; PRT; 181 AA.  
 ID Q917F0;  
 AC 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE CG9027 protein.  
 GN CG9027.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRINBERKELET;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celaliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burlis K.C., Busam M.R., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.G., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,  
 RA Hosten D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy I., Murry D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,  
 RA Relneart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE  
 CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2 SUPEROXIDE + 2 H(+) -> O(2) + H(2)O(2).  
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 DR HSP; P07505; 1SRD.  
 DR FLYBase; FBgn0033631; CG9027.  
 DR InterPro; IPR001424; SOD\_CU\_ZN.  
 DR Pfam; PF00080; SODcu; 1.  
 DR PRINTS; PR00068; CUZNDISMTASE.  
 DR PRODOM; PD000469; SOD\_CU\_ZN; 1.  
 DR PROSITE; PS00087; SOD\_CU\_ZN; 1.  
 DR PROSITE; PS00332; SOD\_CU\_ZN; 2; 1.  
 DR Copper; Oxidoreductase; Zinc.  
 DR SEQUENCE 181 AA; 19221 MW; 68ACE43F8D05CA1 CRC64;  
 Query Match 5.8%; Score 162.5; DB 5; Length 181;  
 Best Local Similarity 31.0%; Pred. No. 2.2e-05;  
 Matches 45; Conservative 25; Mismatches 44; Indels 31; Gaps 7;  
 DB 225 VHGTFIDKQVYG-DLEVSYHLEGFNSDDHDKHLADQIYANGDLTSGCNDICAKYDP- 282  
 43 VKGNVFTQNDCCGVNHVAVQLG-----KEKGKGFHHEKGLTNGCISMGAHNPKVD 96  
 QY 283 -----HEDYHSELGLADIHDDHGVNESHRYSMINIFGDSVLSGSIATHO-RDH 333  
 DB 97 KVDHGGPDHVRH--VGDLGNLEANSIGIIVYITTDVYITLTGKLGITIGRVVYHELED 154  
 QY 334 L---HRSK-----IACVYG 346  
 DB 155 LGLGHNHDSKKTGNAGRIACGVIG 179

Search completed: November 29, 2002, 12:35:07  
 Job time : 95 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 29, 2002, 12:37:00 : Search time 33 Seconds  
(Without alignments)  
2006.835 Million cell updates/sec

Title: US-09-868-760-7

Perfect score: 497  
Sequence: 1 DOBCCNGCQKNDHDDHDDH.....MCGHEVESERVACCVIGRA 497

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A.Geneseq.101002:\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 497   | 100.0       | 497    | 21    | AA193750    |
| 2          | 19    | 3.8         | 19     | 21    | AA193746    |
| 3          | 10    | 2.0         | 10     | 21    | AA193745    |
| 4          | 9     | 1.8         | 9      | 21    | AA193747    |
| 5          | 8     | 1.6         | 18     | 22    | ABBA1196    |
| 6          | 8     | 1.6         | 18     | 22    | ABB25212    |
| 7          | 8     | 1.6         | 18     | 22    | AAAM62056   |
| 8          | 8     | 1.6         | 18     | 22    | AAAM74858   |
| 9          | 8     | 1.6         | 18     | 22    | AAAM34975   |
| 10         | 8     | 1.6         | 18     | 23    | ABGA4634    |

|    |   |     |      |    |          |                    |
|----|---|-----|------|----|----------|--------------------|
| 11 | 8 | 1.6 | 122  | 21 | AA608770 | Arabidopsis thalia |
| 12 | 8 | 1.6 | 122  | 21 | AA643679 | Arabidopsis thalia |
| 13 | 8 | 1.6 | 125  | 21 | AA608769 | Arabidopsis thalia |
| 14 | 8 | 1.6 | 125  | 21 | AA643678 | Arabidopsis thalia |
| 15 | 8 | 1.6 | 259  | 22 | AAU45288 | Arabidopsis thalia |
| 16 | 8 | 1.6 | 474  | 6  | AAU50587 | Arabidopsis thalia |
| 17 | 8 | 1.6 | 476  | 6  | AAU50584 | Arabidopsis thalia |
| 18 | 8 | 1.6 | 477  | 6  | AAU50586 | Arabidopsis thalia |
| 19 | 8 | 1.6 | 1445 | 20 | AAU93505 | Arabidopsis thalia |
| 20 | 8 | 1.6 | 1529 | 7  | AAU93505 | Arabidopsis thalia |
| 21 | 8 | 1.4 | 7    | 21 | AA193746 | Arabidopsis thalia |
| 22 | 7 | 1.4 | 7    | 21 | AA193746 | Arabidopsis thalia |
| 23 | 7 | 1.4 | 34   | 22 | AAU93749 | Arabidopsis thalia |
| 24 | 7 | 1.4 | 58   | 22 | AAU93749 | Arabidopsis thalia |
| 25 | 7 | 1.4 | 66   | 22 | AAU93749 | Arabidopsis thalia |
| 26 | 7 | 1.4 | 108  | 22 | AAU93749 | Arabidopsis thalia |
| 27 | 7 | 1.4 | 125  | 23 | AAU93749 | Arabidopsis thalia |
| 28 | 7 | 1.4 | 127  | 22 | AAU93749 | Arabidopsis thalia |
| 29 | 7 | 1.4 | 172  | 21 | AAU93749 | Arabidopsis thalia |
| 30 | 7 | 1.4 | 172  | 21 | AAU93749 | Arabidopsis thalia |
| 31 | 7 | 1.4 | 192  | 22 | AAU93749 | Arabidopsis thalia |
| 32 | 7 | 1.4 | 205  | 22 | AAU93749 | Arabidopsis thalia |
| 33 | 7 | 1.4 | 222  | 22 | AAU93749 | Arabidopsis thalia |
| 34 | 7 | 1.4 | 222  | 22 | AAU93749 | Arabidopsis thalia |
| 35 | 7 | 1.4 | 225  | 21 | AAU93749 | Arabidopsis thalia |
| 36 | 7 | 1.4 | 226  | 21 | AAU93749 | Arabidopsis thalia |
| 37 | 7 | 1.4 | 226  | 21 | AAU93749 | Arabidopsis thalia |
| 38 | 7 | 1.4 | 227  | 22 | AAU93749 | Arabidopsis thalia |
| 39 | 7 | 1.4 | 244  | 22 | AAU93749 | Arabidopsis thalia |
| 40 | 7 | 1.4 | 253  | 19 | AAU93749 | Arabidopsis thalia |
| 41 | 7 | 1.4 | 255  | 21 | AAU93749 | Arabidopsis thalia |
| 42 | 7 | 1.4 | 259  | 23 | AAU93749 | Arabidopsis thalia |
| 43 | 7 | 1.4 | 260  | 21 | AAU93749 | Arabidopsis thalia |
| 44 | 7 | 1.4 | 268  | 23 | AAU93749 | Arabidopsis thalia |
| 45 | 7 | 1.4 | 275  | 23 | AAU93749 | Arabidopsis thalia |

#### ALIGNMENTS

RESULT 1  
ID AAY93750 standard; Protein: 497 AA.  
XX  
AC AAY93750;  
XX  
DT 03-OCT-2000 (first entry)  
XX  
DE Amino acid sequence of a serine protease inhibitor protein.  
XX  
KW Serine protease inhibitor; green-lipped mussel; anti-thrombin;  
KW divalent metal cation binding activity; dietary supplement;  
KW anticoagulant.  
XX  
OS Perna canaliculus.  
XX  
PN WO200039165-A1.  
XX  
PD 06-JUL-2000.  
XX  
PF 23-DEC-1999; 99WO-NZ00227.  
XX  
PR 23-DEC-1998; 98NZ-0333568.  
XX  
PR 23-JUL-1999; 99NZ-0336906.  
XX  
(HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.  
XX  
PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;  
XX  
WPI: 2000-452375/39.  
XX  
N-PDB: AAA47150. AAA47151.  
XX  
PT New Perna canaliculus serine protease inhibitor protein exhibiting

PT anti-thrombin activity and divalent metal cation binding activity,  
 PR useful as an anticoagulant agent and as a dietary supplement -  
 XX  
 PS Claim 2, Page 2, 44pp: English.  
 XX  
 CC The present sequence represents a serine protease inhibitor  
 CC protein. The protein is isolated from the green-lipped mussel  
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity  
 CC and divalent metal cation binding activity. The serine protease  
 CC inhibitor protein has a molecular weight of about 55 kDa Daltons.  
 CC The protein, and its fragments, are useful in medicaments, in food,  
 CC as dietary supplements or as flocculation agents. In the dietary  
 CC supplements, the protein is associated with or bound to at least one  
 CC divalent cation (such as calcium, magnesium or zinc) of dietary  
 CC significance. The proteins or their fragments are also useful as  
 CC anticoagulant agents.  
 CC  
 XX  
 XX Sequence 497 AA:

[illegible]

XX Perna canaliculus.  
OS  
XX  
XX PN WO2000039165-A1.  
XX  
XX PD 06-JUL-2000.  
XX  
XX PF 23-DEC-1999; 99WO-NZ00227.  
XX  
XX PR 23-DEC-1998; 98NZ-0333568.  
XX PR 23-JUL-1999; 99NZ-0336906.  
XX  
XX PA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.  
PA  
XX PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;  
XX WPI; 2000-452375/39.  
XX  
XX PT New Perna canaliculus serine protease inhibitor protein exhibiting  
XX PT anti-thrombin activity and divalent metal cation binding activity,  
XX PT useful as an anticoagulant agent and as a dietary supplement -  
XX  
XX SS Claim 1; Page 16; 44pp; English.

|  |          |   |
|--|----------|---|
|  | CC       | AAV93745-49 represent peptides derived from a serine protease inhibitor |
|  | CC       | protein. The protein is isolated from the green-lipped mussel           |
|  | CC       | (Perna canaliculus), and exhibits inter alia, anti-thrombin activity    |
|  | CC       | and divalent metal cation binding activity. The serine protease         |
|  | CC       | inhibitor protein has a molecular weight of about 55 kilo Daltons.      |
|  | CC       | The protein, and its fragments, are useful in medicaments, in food,     |
|  | CC       | as dietary supplements or as bioremediation agents. In the dietary      |
|  | CC       | supplements, the protein is associated with or bound to at least one    |
|  | CC       | divalent cation (such as calcium, magnesium or zinc) of dietary         |
|  | CC       | significance. The proteins or their fragments are also useful as        |
|  | CC       | anticoagulant agents.   |
|  | XX       |   |
|  | XX       | Sequence 19 AA:   |
|  | QY       | Query Match 3.8%; Score 19; DB 21; Length 19;                           |
|  | Dd       | Best Local Similarity 100.0%; Pred. No. 1,8e-11;                        |
|  |          | Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;             |
|  | OY       | 479 OGGHEVESERVACCVIGRA 497<br>   |
|  | Dd       | 1 OGGHEVESERVACCVIGRA 19  |
|  | RESULT 3 |   |
|  | ID       | AAV93745  |
|  | XX       | AAV93745 standard; peptide; 10 AA.                                      |
|  | AC       | AAV93745:   |
|  | XX       |   |
|  | DT       | 03-OCT-2000 (first entry)   |
|  | XX       |   |
|  | DE       | Peptide derived from a serine protease inhibitor protein.               |
|  | XX       |   |
|  | KW       | Serine protease inhibitor; green-lipped mussel; anti-thrombin;          |
|  | KW       | divalent metal cation binding activity; dietary supplement;             |
|  | XX       | anticoagulant.  |
|  | OS       | Perna canaliculus.  |
|  | XX       |   |
|  | FN       | WO200039165-A1.   |
|  | FD       | 06-JUL-2000.  |
|  | XX       |   |
|  | FE       | 23-DEC-1999; 99WO-NZ00227.  |
|  | XX       |   |
|  | ER       | 23-DEC-1998; 98NZ-0333568.  |
|  | XX       |   |
|  | XX       | 23-JUL-1999; 99NZ-0336906.  |
|  | XX       |   |
|  | XA       | (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.                       |

PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;  
 XX  
 DR WPI: 2000-452375/39.  
 XX  
 PT New Perna canaliculus serine protease inhibitor protein exhibiting  
 PT anti-thrombin activity and divalent metal cation binding activity,  
 PT useful as an anticoagulant agent and as a dietary supplement -  
 XX  
 PS Claim 1: Page 16; 44pp; English.  
 CC  
 CC AAY93745-49 represent peptides derived from a serine protease inhibitor  
 CC protein. The protein is isolated from the green-lipped mussel  
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity  
 CC and divalent metal cation binding activity. The serine protease  
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.  
 CC The protein, and its fragments, are useful in medicaments, in food,  
 CC as dietary supplements or as bioremediation agents. In the dietary  
 CC supplements, the protein is associated with or bound to at least one  
 CC divalent cation (such as calcium, magnesium or zinc) of dietary  
 CC significance. The proteins or their fragments are also useful as  
 CC anticoagulant agents.  
 CC  
 SQ Sequence 10 AA;  
 Query Match 2.0%; Score 10; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.011;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 DGECCNDGON 10  
 Db 1 DGECCNDGON 10  
 RESULT 4  
 AAY93747  
 ID AAY93747 standard; peptide: 9 AA.  
 AC AAY93747:  
 DT 03-OCT-2000 (first entry)  
 DE Peptide derived from a serine protease inhibitor protein.  
 XX  
 XX Serine protease inhibitor; green-lipped mussel; anti-thrombin;  
 XX di-valent metal cation binding activity; dietary supplement;  
 XX anticoagulant.  
 XX  
 XX Perna canaliculus.  
 OS WO200039165-A1.  
 PN 06-JUL-2000.  
 PD 23-DEC-1999; 99WO-NZ00227.  
 PF 23-DEC-1998; 98NZ-0333568.  
 PR 23-JUL-1999; 99NZ-0336906.  
 XX  
 XX (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.  
 PA  
 XX Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;  
 PI WPI: 2000-452375/39.  
 DR  
 XX  
 XX New Perna canaliculus serine protease inhibitor protein exhibiting  
 PT anti-thrombin activity and divalent metal cation binding activity,  
 PT useful as an anticoagulant agent and as a dietary supplement -  
 XX  
 PS Claim 1: Page 16; 44pp; English.  
 CC  
 CC AAY93745-49 represent peptides derived from a serine protease inhibitor  
 CC protein. The protein is isolated from the green-lipped mussel  
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity  
 CC and divalent metal cation binding activity. The serine protease  
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.  
 CC The protein, and its fragments, are useful in medicaments, in food,  
 CC as dietary supplements or as bioremediation agents. In the dietary  
 CC supplements, the protein is associated with or bound to at least one  
 CC divalent cation (such as calcium, magnesium or zinc) of dietary  
 CC significance. The proteins or their fragments are also useful as  
 CC anticoagulant agents.  
 CC  
 SQ Sequence 10 AA;  
 Query Match 2.0%; Score 10; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.011;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 DGECCNDGON 10  
 Db 1 DGECCNDGON 10

```

CC (Perna canaliculus) , and exhibits, inter alia, anti-thrombin activity
CC and divalent metal cation binding activity. The serine protease
CC inhibitor protein has a molecular weight of about 55 kilo daltons.
CC The protein, and its fragments, are useful in medicaments, in food,
CC as dietary supplements or as bioremediation agents. In the dietary
CC supplements, the protein is associated with or bound to at least one
CC divalent cation (such as calcium, magnesium or zinc) of dietary
CC significance. The proteins or their fragments are also useful as
CC anticoagulant agents.
CC
XX
XX Sequence 9 AA:
XX
XX
XX Query Match 1.8%; Score 9; DB 21; Length 9;
XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 348 GQSHPEIVH 356
XX 11111111
XX 1 GQSHPEIVH 9
XX
XX RESULT 5
XX ID ABB41196 standard; Peptide; 18 AA.
XX ABB41196
XX AC ABB41196;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #8702 encoded by human foetal liver single exon probe.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
XX Homo sapiens.
XX
XX OS WO200157277-A2.
XX PM
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00669.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 33831; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 18 AA:
XX

```

Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 HDDHDDD 30  
          |||||  
DB 9 HDDHDDD 16

RESULT 6  
ABR25212  
ID ABR25212 standard; Protein: 18 AA.  
XX  
AC ABR25212;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #7211 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human: gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001WO-US00666.  
XX  
PF 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID No 26982; 530pp; English.  
XX  
SQ The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABR21535-ABR41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 18 AA:

Query Match 1.6%; Score 8; DB 22; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 HDDHDDD 30  
          |||||  
DB 9 HDDHDDD 16

RESULT 7  
AAM62056  
ID AAM62056 standard; Protein: 18 AA.  
XX  
AC AAM62056;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34161.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001WO-US00667.  
XX  
PF 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
PS Example 4; SEQ ID NO: 34161; 650pp + Sequence Listing; English.  
XX  
SQ The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
SQ Sequence 18 AA:

Query Match 1.6%; Score 8; DB 22; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 HDDHDDD 30  
          |||||  
DB 9 HDDHDDD 16

RESULT 8  
AAM74858  
ID AAM74858 standard; Protein: 18 AA.  
XX  
AC AAM74858;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 35164.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukemia; lymphoma; myeloma.  
XX

```
OS Homo sapiens.
XX
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 35164; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX
XX Sequence 18 AA:
SQ
Query Match 1.6%; Score 8; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 23 HDDHDDDD 30
DB 9 HDDHDDDD 16
RESULT 9
AAM34975
ID AAM34975 standard; Protein; 18 AA.
XX
XX AAM34975;
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #9012 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
```

```
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID NO 35244; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see A131315-A157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX
XX Sequence 18 AA:
SQ
Query Match 1.6%; Score 8; DB 22; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 23 HDDHDDDD 30
DB 9 HDDHDDDD 16
RESULT 10
ABG44634
ID ABG44634 standard; Peptide; 18 AA.
XX
XX ABG44634;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human peptide encoded by genome-derived single exon probe SEQ ID 34299.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX
XX Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX
XX 26-MAY-2000; 2000US-207456P.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-234687P.
XX
XX 27-SEP-2000; 2000US-236359P.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
```

PT measure gene expression in human lung samples -  
XX  
PS Claim 27; SEQ ID No 34299; 634bp; English.  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes, the novel set of probes which hybridize at high stringency to a  
CC nucleic acid expressed in the human lung, measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray, assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC hemangioendothelioma, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPD at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
SQ Sequence 18 AA:  
Query Match 1.6%; Score 8; DB 23; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 HDHDDDD 30  
          |||||||  
Db      9 HDHDDDD 16  
  
RESULT 11  
AAG08770  
ID AAG08770 standard; Protein: 122 AA.  
XX  
XX AAG08770:  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6439.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
FN EPI033405-A2.

XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000: 2000EP-0301439.  
XX  
XX 25-FEB-1999: 99US-0121825.  
PR 05-MAR-1999: 99US-0123180.  
PR 09-MAR-1999: 99US-0123548.  
PR 23-MAR-1999: 99US-0125788.  
PR 25-MAR-1999: 99US-0126264.  
PR 01-APR-1999: 99US-0126785.  
PR 06-APR-1999: 99US-0126234.  
PR 08-APR-1999: 99US-0126714.  
PR 16-APR-1999: 99US-0126845.  
PR 19-APR-1999: 99US-0130077.  
PR 21-APR-1999: 99US-0130449.  
PR 23-APR-1999: 99US-0130510.  
PR 28-APR-1999: 99US-0130891.  
PR 30-APR-1999: 99US-0131449.  
PR 30-APR-1999: 99US-0132048.  
PR 04-MAY-1999: 99US-0132407.  
PR 05-MAY-1999: 99US-0132484.  
PR 06-MAY-1999: 99US-0132485.  
PR 06-MAY-1999: 99US-0132486.  
PR 07-MAY-1999: 99US-0132487.  
PR 11-MAY-1999: 99US-0132863.  
PR 14-MAY-1999: 99US-0134256.  
PR 14-MAY-1999: 99US-0134218.  
PR 14-MAY-1999: 99US-0134219.  
PR 14-MAY-1999: 99US-0134221.  
PR 18-MAY-1999: 99US-0134370.  
PR 19-MAY-1999: 99US-0134768.  
PR 20-MAY-1999: 99US-0134941.  
PR 21-MAY-1999: 99US-0135124.  
PR 24-MAY-1999: 99US-0135353.  
PR 25-MAY-1999: 99US-0135629.  
PR 27-MAY-1999: 99US-0136021.  
PR 28-MAY-1999: 99US-0136392.  
PR 01-JUN-1999: 99US-0136782.  
PR 03-JUN-1999: 99US-0137222.  
PR 04-JUN-1999: 99US-0137528.  
PR 07-JUN-1999: 99US-0137502.  
PR 08-JUN-1999: 99US-0137724.  
PR 10-JUN-1999: 99US-0138094.  
PR 10-JUN-1999: 99US-0138540.  
PR 14-JUN-1999: 99US-0138847.  
PR 16-JUN-1999: 99US-0139119.  
PR 16-JUN-1999: 99US-0139452.  
PR 17-JUN-1999: 99US-0139453.  
PR 18-JUN-1999: 99US-0139492.  
PR 18-JUN-1999: 99US-0139454.  
PR 18-JUN-1999: 99US-0139455.  
PR 18-JUN-1999: 99US-0139456.  
PR 18-JUN-1999: 99US-0139457.  
PR 18-JUN-1999: 99US-0139458.  
PR 18-JUN-1999: 99US-0139459.  
PR 18-JUN-1999: 99US-0139460.  
PR 18-JUN-1999: 99US-0139461.  
PR 18-JUN-1999: 99US-0139462.  
PR 18-JUN-1999: 99US-0139463.  
PR 18-JUN-1999: 99US-0139750.  
PR 18-JUN-1999: 99US-0139763.  
PR 21-JUN-1999: 99US-0139817.  
PR 22-JUN-1999: 99US-0139899.  
PR 23-JUN-1999: 99US-0140353.  
PR 24-JUN-1999: 99US-0140354.  
PR 28-JUN-1999: 99US-0140655.  
PR 29-JUN-1999: 99US-0140682.  
PR 30-JUN-1999: 99US-0140991.  
PR 01-JUL-1999: 99US-0141287.  
PR 01-JUL-1999: 99US-0141842.  
PR 01-JUL-1999: 99US-0142154.



```

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 23-JUL-1999; 99US-0145142.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145918.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 04-AUG-1999; 99US-0146388.
PR 04-AUG-1999; 99US-0147038.
PR 05-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 30-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153738.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 07-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 25-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 8; DB 21; Length 122;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 402 EDLSHNRH 409
Db 76 EDLSHNRH 83

RESULT 12
ID AG43679 standard; Protein; 122 AA.
XX AG43679;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 54622.
XX Arabidopsis thaliana protein fragment SEQ ID NO: 54622.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX 06-SEP-2000.
PD 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-012180.

```

|    |             |              |
|----|-------------|--------------|
| PR | 23-MAR-1999 | 9905-0125348 |
| PR | 03-MAR-1999 | 9905-0125788 |
| PR | 25-MAR-1999 | 9905-0126264 |
| PR | 25-MAR-1999 | 9905-0126785 |
| PR | 01-MAR-1999 | 9905-0127442 |
| PR | 06-APR-1999 | 9905-0128824 |
| PR | 16-APR-1999 | 9905-0128874 |
| PR | 19-APR-1999 | 9905-0130077 |
| PR | 23-APR-1999 | 9905-0130449 |
| PR | 23-APR-1999 | 9905-0130851 |
| PR | 26-APR-1999 | 9905-0131444 |
| PR | 30-APR-1999 | 9905-0131404 |
| PR | 04-MAY-1999 | 9905-0133407 |
| PR | 05-MAY-1999 | 9905-0133488 |
| PR | 06-MAY-1999 | 9905-0133486 |
| PR | 07-MAY-1999 | 9905-0132863 |
| PR | 11-MAY-1999 | 9905-0134256 |
| PR | 14-MAY-1999 | 9905-0134216 |
| PR | 14-MAY-1999 | 9905-0134215 |
| PR | 14-MAY-1999 | 9905-0134271 |
| PR | 18-MAY-1999 | 9905-0134768 |
| PR | 19-MAY-1999 | 9905-0134841 |
| PR | 20-MAY-1999 | 9905-0135124 |
| PR | 21-MAY-1999 | 9905-0135263 |
| PR | 22-MAY-1999 | 9905-0136021 |
| PR | 23-MAY-1999 | 9905-0136782 |
| PR | 28-MAY-1999 | 9905-0136782 |
| PR | 01-JUN-1999 | 9905-0137522 |
| PR | 03-JUN-1999 | 9905-0137528 |
| PR | 07-JUN-1999 | 9905-0137724 |
| PR | 08-JUN-1999 | 9905-0138094 |
| PR | 10-JUN-1999 | 9905-0138540 |
| PR | 10-JUN-1999 | 9905-0138847 |
| PR | 16-JUN-1999 | 9905-0139119 |
| PR | 16-JUN-1999 | 9905-0139452 |
| PR | 16-JUN-1999 | 9905-0139453 |
| PR | 16-JUN-1999 | 9905-0139452 |
| PR | 18-JUN-1999 | 9905-0139455 |
| PR | 18-JUN-1999 | 9905-0139456 |
| PR | 18-JUN-1999 | 9905-0139458 |
| PR | 18-JUN-1999 | 9905-0139457 |
| PR | 18-JUN-1999 | 9905-0139458 |
| PR | 18-JUN-1999 | 9905-0139460 |
| PR | 18-JUN-1999 | 9905-0139461 |
| PR | 18-JUN-1999 | 9905-0139462 |
| PR | 18-JUN-1999 | 9905-0139463 |
| PR | 18-JUN-1999 | 9905-0139470 |
| PR | 18-JUN-1999 | 9905-0139763 |
| PR | 21-JUN-1999 | 9905-0139817 |
| PR | 22-JUN-1999 | 9905-0140355 |
| PR | 23-JUN-1999 | 9905-0140353 |
| PR | 24-JUN-1999 | 9905-0140354 |
| PR | 24-JUN-1999 | 9905-0140693 |
| PR | 26-JUN-1999 | 9905-0140823 |
| PR | 30-JUN-1999 | 9905-0140991 |
| PR | 01-JUL-1999 | 9905-0141287 |
| PR | 01-JUL-1999 | 9905-0141484 |
| PR | 02-JUL-1999 | 9905-0142124 |
| PR | 06-JUL-1999 | 9905-0142053 |
| PR | 08-JUL-1999 | 9905-0142803 |
| PR | 09-JUL-1999 | 9905-0142803 |
| PR | 12-JUL-1999 | 9905-0142877 |
| PR | 13-JUL-1999 | 9905-0143524 |
| PR | 14-JUL-1999 | 9905-0143524 |

[illegible]



PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144332.  
 PR 20-JUL-1999; 99US-0144384.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145112.  
 PR 23-JUL-1999; 99US-0145118.  
 PR 23-JUL-1999; 99US-0145274.  
 PR 26-JUL-1999; 99US-0145273.  
 PR 27-JUL-1999; 99US-0145318.  
 PR 27-JUL-1999; 99US-0145319.  
 PR 28-JUL-1999; 99US-0145385.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147303.  
 PR 05-AUG-1999; 99US-0147304.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 06-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149802.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 23-AUG-1999; 99US-0150386.  
 PR 25-AUG-1999; 99US-0150386.  
 PR 26-AUG-1999; 99US-0150684.  
 PR 27-AUG-1999; 99US-0151062.  
 PR 27-AUG-1999; 99US-0151063.  
 PR 27-AUG-1999; 99US-0151064.  
 PR 30-AUG-1999; 99US-0151293.  
 PR 31-AUG-1999; 99US-0151498.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155658.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0158029.  
 PR 07-OCT-1999; 99US-0158265.  
 PR 08-OCT-1999; 99US-0158322.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 21-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 22-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.6%; Score 8; DB 21; Length 125;  
 Best Local Similarity 100.0%; Pred. No 9; 4; Mismatches 0;  
 Matches 8; Conservative 0; Indels 0; Caps 0;

OY 402 EDLSHRH 409  
 DB 79 EDLSHRH 86

RESULT 14  
 AAC43678 standard; Protein: 125 AA.  
 ID AAC43678  
 XX AAC43678;  
 AC  
 XX 18-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 54621.  
 XX  
 DE Protein identification: signal transduction pathway; metabolic pathway;  
 KM hydrolisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.  
 OS Arabidopsis thaliana.  
 OS  
 PN EPI033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 03-MAR-1999; 99US-0123180.  
 PR 23-MAR-1999; 99US-0123548.  
 PR 25-MAR-1999; 99US-0125788.  
 PR 26-MAR-1999; 99US-0126264.  
 PR 01-APR-1999; 99US-0126785.  
 PR 06-APR-1999; 99US-0127462.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132488.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134220.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134911.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140683.  
PR 28-JUN-1999; 99US-0140683.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144375.  
PR 19-JUL-1999; 99US-0144376.  
PR 19-JUL-1999; 99US-0144377.  
PR 19-JUL-1999; 99US-0144378.  
PR 19-JUL-1999; 99US-0144379.  
PR 19-JUL-1999; 99US-0144380.  
PR 19-JUL-1999; 99US-0144381.  
PR 19-JUL-1999; 99US-0144382.  
PR 19-JUL-1999; 99US-0144383.  
PR 19-JUL-1999; 99US-0144384.  
PR 20-JUL-1999; 99US-0144385.  
PR 20-JUL-1999; 99US-0144386.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148321.  
PR 13-AUG-1999; 99US-0148365.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 18-AUG-1999; 99US-0149175.  
PR 20-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157553.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158380.  
PR 12-OCT-1999; 99US-0158383.  
PR 13-OCT-1999; 99US-0159283.  
PR 13-OCT-1999; 99US-0159284.  
PR 14-OCT-1999; 99US-0159285.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.

```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 1.68; Score 8; DB 21; Length 125;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 402 EDLSHHRH 409
DB 79 EDLSHHRH 86

```

```

RESULT 15
AAU45288
ID AAU45288 standard; Protein: 259 AA.

```

```

XX AAU45288;

```

```

XX 27-FEB-2002 (first entry)

```

```

DE Propionibacterium acnes immunogenic protein #6184.

```

```

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

```

```

XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

```

```

XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

```

```

XX dermatological; osteopathic; neuroprotectant.

```

```

XX Propionibacterium acnes.

```

```

XX WO200181581-A2.

```

```

XX 01-NOV-2001.

```

```

XX 20-APR-2001; 2001WO-US12865.

```

```

XX 21-APR-2000; 2000US-199047P.

```

```

XX 02-JUN-2000; 2000US-208841P.

```

```

XX 07-JUL-2000; 2000US-216747P.

```

```

XX (CORI-) CORIXA CORP.

```

```

XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;

```

```

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

```

```

XX WPI; 2001-616774/71.

```

```

XX N-PSDB; AAS59525.

```

```

XX Propionibacterium acnes polypeptides and nucleic acids useful for

```

```

XX PT vaccinating against and diagnosing infections, especially useful for

```

```

XX PT treating acne vulgaris -

```

```

XX Example 1: SEQ ID No 6483; 1069PP; English.

```

```

XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic

```

```

XX polypeptides. The proteins and their associated DNA sequences are used in

```

```

XX the treatment, prevention and diagnosis of medical conditions caused by

```

```

CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

```

```

XX Sequence 259 AA;

```

```

SQ

```

```

Query Match 1.68; Score 8; DB 22; Length 259;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 121 DLGDLVDD 128
DB 105 DLGDLVDD 112

```

```

Search completed: November 29, 2002, 12:44:41
Job time : 37 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 07:50:58 ; Search time 2055 seconds

(without alignments)  
11750.594 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gaggggggagcagtgtaacga.....gctgtgtatagagcggca 1491

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gp\_est1:\*  
10: gp\_est2:\*  
11: gp\_hic:\*  
12: gp\_est3:\*  
13: gp\_est4:\*  
14: gp\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gp\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pin:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB | ID       | Description         |
|------------|-------|-------|--------|----|----------|---------------------|
| 1          | 65.6  | 4.4   | 908    | 17 | A2548467 | A2548467 ENTKE30TR  |
| 2          | 64.2  | 4.3   | 877    | 17 | A2531291 | A2531291 ENTBO34TR  |
| 3          | 64.2  | 4.3   | 912    | 17 | A2551092 | A2551092 ENTREJ22TR |
| 4          | 63.4  | 4.3   | 468    | 12 | BG553005 | BG553005 ddb82e09.  |
| 5          | 62.2  | 4.2   | 795    | 17 | A2528485 | A2528485 ENTOM64TR  |
| 6          | 62.2  | 4.2   | 823    | 17 | A2676218 | A2676218 ENTKE36TR  |

|    |      |     |     |    |          |                     |
|----|------|-----|-----|----|----------|---------------------|
| 7  | 62.2 | 4.2 | 843 | 17 | BH139532 | BH139532 ENTNG88TR  |
| 8  | 58.8 | 3.9 | 740 | 10 | AV679268 | AV679268 A679268R   |
| 9  | 58.4 | 3.9 | 976 | 17 | BH149983 | BH149983 ENTOD93TR  |
| 10 | 57.2 | 3.8 | 843 | 17 | A2531618 | A2531618 ENTDV54TR  |
| 11 | 57.2 | 3.8 | 887 | 14 | BQ731479 | BQ731479 AGENCOURT  |
| 12 | 56.8 | 3.8 | 825 | 13 | BJ374097 | BJ374097 BQ374097   |
| 13 | 56.6 | 3.8 | 644 | 14 | BQ834753 | BQ834753 PO_ad_03A  |
| 14 | 56.6 | 3.8 | 661 | 14 | BQ834718 | BQ834718 PO_ad_02F  |
| 15 | 56.4 | 3.8 | 677 | 13 | BH161314 | BH161314 EST563837  |
| 16 | 56   | 3.8 | 880 | 17 | A2669474 | A2669474 ENTIR88TR  |
| 17 | 55.6 | 3.7 | 742 | 10 | AV674403 | AV674403 A674403R   |
| 18 | 55.4 | 3.7 | 607 | 13 | BJ348812 | BJ348812 BJT48812   |
| 19 | 55.2 | 3.7 | 900 | 17 | A2549980 | A2549980 ENTDD94TR  |
| 20 | 55   | 3.7 | 906 | 17 | BH153606 | BH153606 ENTTS83TR  |
| 21 | 54.8 | 3.7 | 816 | 17 | A2535744 | A2535744 ENTQ25TR   |
| 22 | 54.8 | 3.7 | 891 | 17 | A2683582 | A2683582 ENTKR47TR  |
| 23 | 54.6 | 3.7 | 825 | 13 | BJ374097 | BJ374097 BQ374097   |
| 24 | 54   | 3.6 | 751 | 13 | BJ374223 | BJ374223 BQ374223   |
| 25 | 53.2 | 3.6 | 890 | 17 | BH146886 | BH146886 ENTPK48TR  |
| 26 | 52.8 | 3.5 | 493 | 13 | BH170559 | BH170559 EST573082  |
| 27 | 52.8 | 3.5 | 880 | 17 | A2529191 | A2529191 ENTBOV68TR |
| 28 | 52.6 | 3.5 | 931 | 17 | BH160272 | BH160272 ENTBOV49TR |
| 29 | 52.4 | 3.5 | 890 | 17 | A2530768 | A2530768 ENTBH54TR  |
| 30 | 52.2 | 3.5 | 548 | 13 | BH159818 | BH159818 EST562341  |
| 31 | 52   | 3.5 | 849 | 17 | A2546009 | A2546009 ENTPE53TR  |
| 32 | 51.8 | 3.5 | 445 | 17 | BH167469 | BH167469 EST568932  |
| 33 | 51.8 | 3.5 | 723 | 13 | BQ339844 | BQ339844 BQ339844   |
| 34 | 51.2 | 3.4 | 598 | 13 | BH170666 | BH170666 EST573189  |
| 35 | 51.2 | 3.4 | 605 | 13 | BH163520 | BH163520 EST566043  |
| 36 | 51.2 | 3.4 | 645 | 13 | BH165350 | BH165350 EST567873  |
| 37 | 51.2 | 3.4 | 647 | 13 | BH163120 | BH163120 EST565643  |
| 38 | 51.2 | 3.4 | 671 | 13 | BH160252 | BH160252 EST562775  |
| 39 | 51.2 | 3.4 | 694 | 14 | BH168242 | BH168242 EST570765  |
| 40 | 51.2 | 3.4 | 717 | 13 | BH160500 | BH160500 EST5653023 |
| 41 | 51.2 | 3.4 | 747 | 13 | BH162732 | BH162732 EST565255  |
| 42 | 51.2 | 3.4 | 756 | 13 | BH162492 | BH162492 EST565015  |
| 43 | 51.2 | 3.4 | 764 | 13 | BJ339827 | BJ339827 BQ339827   |
| 44 | 51.2 | 3.4 | 794 | 13 | BH159906 | BH159906 EST562429  |
| 45 | 50.8 | 3.4 | 323 | 13 | BH171402 | BH171402 EST573925  |

#### ALIGNMENTS

RESULT 1  
A2548467  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

A2548467 908 bp DNA linear GSS 14-NOV-2000  
ENTKE30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.  
A2548467  
A2548467.1 GI:11172102  
GSS.  
Entamoeba histolytica.  
Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.  
1 (bases 1 to 908)  
Determination of clone end sequences from Entamoeba histolytica  
Loftus, B., Van Aken, S. and Fraser, C.  
HM1:IMSS sheared DNA library  
Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b1loftus@igf.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: Shotgun  
High quality sequence start: 17  
High quality sequence stop: 828.

## FEATURES

source

location/Qualifiers  
1..908  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone.lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOSt. Site 1: Best I: Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 434 a 16 c 289 g 169 t

## ORIGIN

Query Match 4.4%; Score 65.6; DB 17; Length 908;  
Best Local Similarity 43.8%; Pred. No. 1e-06; Mismatches 429; Indels 3; Gaps 1;  
Matches 336; Conservative 0;

17 ACAGTGGGAGAGCAAGAGTACACGACCCAGCAGATGATCAGTACAGC 76  
20 AAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 79  
77 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 136  
80 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 139  
137 TGCTGAGCAGCCTGACCCAGTGCATGAGCAGTACAGTGTGACAG 196  
140 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 199  
197 GAGCTGTTATCTAGAACTTCCTTGTGATCAACAGTGAAGCATGAG 256  
200 AAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 259  
257 ACCATCTGATGATGATGATGATGATGATGATGATGATGATGATG 316  
260 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 319  
317 GCGAAGTGTACATGCTACCCAGAAAGATGCTGACCTGCTGCTGG 376  
320 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 379  
377 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 436  
380 AAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 439  
437 GTACAGCAGCAACAGCAGAGCTCTGATGACACTCAATATTTACA 496  
440 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 499  
497 ACACCGATGCTGATACCCAGCAGTGAATGCTGCTGCTGCTGCTG 556  
500 ATGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATG 559  
557 CTGCGCCAGCAACAGCAGCT--GCTCTACATCAGAGCTGAGAGA 613  
560 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 619  
614 ATTATGCCATGCTGATGATGATGATGATGATGATGATGATGATG 673  
620 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 679  
674 TCGAGGAGCAATGATTTCAACAGTGTGATGATGATGATGATGAT 733  
680 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 739

QY 734 TAGAGGATTTAATGATGATGATGATGATGATGATGATGATGATG 781  
Db 740 ATGATGATGATGATGATGATGATGATGATGATGATGATGATG 787

## FEATURES

source

RESULT 2  
A2531291/c 877 bp DNA linear GSS 03-NOV-2000  
LOCUS ENT080344R Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic DNA sequence.  
ACCESSION A2531291  
VERSION A2531291.1 GI:11085838  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 877)  
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3543  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Names are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 22  
High quality sequence stop: 829.  
Location/Qualifiers  
1..877  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone.lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOSt. Site 1: Best I: Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a light size distribution (~2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

## BASE COUNT

156 a 268 c 29 g 424 t

## ORIGIN

Query Match 4.3%; Score 64.2; DB 17; Length 877;  
Best Local Similarity 43.7%; Pred. No. 2.4e-06; Mismatches 354; Indels 0; Gaps 0;  
Matches 276; Conservative 1;

QY 1 GAYGGGAGCAGTGTACGATGGCGAGAACAGATGACCCAGTGAAGCAG 60  
Db 776 GATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
QY 61 GATCAGCAGTGAAGCAGATGATGATGATGATGATGATGATGATG 120  
Db 716 GAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 657  
QY 121 GAACCAACCCATGATGCTGACGCTTCAACAGTGTGATGATGATGAT 180  
Db 656 GAGAGAGAGATGATGATGATGATGATGATGATGATGATGATG 597



[illegible]

method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is detailed in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

REFERENCE 1 (bases 1 to 468)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strusberg, Ph.D.  
Email: cgaaps-remail.nih.gov  
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: Xenopus clones from this library are available  
through the I.M.A.G.E. Consortium/LNL at: [InfoImage.lnl.gov](mailto:InfoImage.lnl.gov)  
Seq primer: -40RP from Glibco  
High quality sequence stop: 404.

FEATURES  
source  
1..468  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone\_image="4203736"  
/clone\_lib="NICHD XGC Emb4"  
/dev\_stage="embryo, stage 31-32"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: whole embryo; Vector: PCMV-SF0RT6; Site:1;  
NotI; Site:2; SalI; Cloned unidirectionally. Primer: Oligo  
dT. Average insert size 2.1 kb. Constructed by Life  
Technologies. Note: This is a Xenopus Gene Collection (XGC  
) library."

BASE COUNT 164 a 66 c 144 g 94 t

Query Match 4.3%; Score 63.4; DB 12; Length 468;  
Best Local Similarity 48.0%; Pred. No. 2.4e-06;  
Matches 178; Conservative 1; Mismatches 192; Indels 0; Gaps 0;

OY 1 GAYGGGAGAGTGTACGATGGCGAGACAAGATGACCCACCATGACGACACACGAT 60  
DB 63 GATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 122  
OY 61 GATGACCATGCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
DB 123 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 182  
OY 121 GAACCAACCCGTATATGCTAGACACCTTCCACACGATGCGACGATGATGATG 180  
DB 183 GATACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 242  
OY 181 TCACAGAAAGGCTGATGAGCTGTTTATCTAGAACTTCATCTTTCGATTCACACA 240  
DB 243 GATACGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302  
OY 241 GAAGACCATGACGACCATCATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
DB 303 GATACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 362  
OY 301 GGTGTGATTTCTATTGGCGAACTGTACATGCTCAACCCAGAAAAACATGCTGCT 360  
DB 363 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 422  
OY 361 GACCTCGGTGA 371  
DB 423 GATGATGATGA 433

RESULT 5  
A2528485 795 bp DNA linear GSS 03-NOV-2000  
LOCUS ENTMOEBA histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic DNA sequence.  
ACCESSION A2528485  
VERSION A2528485.1 GI:11080747  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica  
REFERENCE 1 (bases 1 to 795)  
AUTHORS Loftus, B., Van Aken, S., and Fraser, C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3343  
Email: [bjloftus@igf.org](mailto:bjloftus@igf.org)  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: Shotgun  
High quality sequence start: 23  
High quality sequence stop: 790.  
Location/Qualifiers  
1..795  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHD1; Site:1; Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Bartell, Oxford University Press, 1999)."

BASE COUNT 319 a 72 c 175 g 229 t

Query Match 4.2%; Score 62.2; DB 17; Length 795;  
Best Local Similarity 47.5%; Pred. No. 7.9e-06;  
Matches 184; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

OY 32 AAGTGGACCCAGTGGACGACACGATGATGATGATGATGATGATGATGATGATGATGAT 91  
DB 317 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 376  
OY 92 AAGCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 151  
DB 377 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 436  
OY 152 ACCACCATGTCATGACGACATGATGATGATGATGATGATGATGATGATGATGATGAT 211  
DB 437 ATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 496  
OY 212 AACTTCATCTTGTGATTCACACCAAGTGAAGACATGACGACGACCATCATGATGAT 271  
DB 497 ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 556  
OY 272 ATCTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 331  
DB 557 ATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 616  
OY 332 CTCACCCGAAAAACATGCTGACCTGGGACCTGGGACCTGGTGGATGATGATGATGAT 391  
DB 617 ATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 676  
OY 392 GCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 418  
DB 677 ATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 703

| RESULT 6              | 42676218/c   | LOCUS                  | 823 bp         | DNA                   | linear | GSS 14-DEC-2000 |
|-----------------------|--|------------------------|----------------|-----------------------|--------|-----------------|
| DEFINITION            | ENTRXP36TR   | Entamoeba histolytica  | Sheared DNA    | Entamoeba histolytica |        |                 |
| ACCESSION             | 42676218   | genomic, DNA sequence. |                |                       |        |                 |
| VERSION               | 42676218.1   | GI:11813364            |                |                       |        |                 |
| KEYWORDS              | GSS.   |                        |                |                       |        |                 |
| SOURCE                | Entamoeba histolytica.   |                        |                |                       |        |                 |
| ORGANISM              | Entamoeba histolytica  |                        |                |                       |        |                 |
| REFERENCE             | Eukaryota; Entamoebidae; Entamoeba.                                |                        |                |                       |        |                 |
| AUTHORS               | 1 (bases 1 to 823)   |                        |                |                       |        |                 |
| TITLE                 | Loftus,B., Van Aken,S. and Fraser,C.                               |                        |                |                       |        |                 |
|                       | Determination of clone end sequences from Entamoeba histolytica    |                        |                |                       |        |                 |
|                       | HML:IMSS sheared DNA library                                       |                        |                |                       |        |                 |
| JOURNAL               | Unpublished (2000)   |                        |                |                       |        |                 |
| COMMENT               | Contact: Brendan J Loftus  |                        |                |                       |        |                 |
|                       | Department of Eukaryotic Genomics                                  |                        |                |                       |        |                 |
|                       | The Institute for Genomic Research                                 |                        |                |                       |        |                 |
|                       | 9712 Medical Center Dr., Rockville, MD 20850, USA                  |                        |                |                       |        |                 |
|                       | Tel: 301 838 0208  |                        |                |                       |        |                 |
|                       | Fax: 301 838 3543  |                        |                |                       |        |                 |
|                       | Email: b.loftus@igr.org  |                        |                |                       |        |                 |
|                       | Clones are derived from the Entamoeba histolytica HML:IMSS sheared |                        |                |                       |        |                 |
|                       | DNA library  |                        |                |                       |        |                 |
|                       | Seq primer: M13-Reverse  |                        |                |                       |        |                 |
|                       | Class: Shotgun   |                        |                |                       |        |                 |
|                       | High quality sequence start: 17                                    |                        |                |                       |        |                 |
|                       | High quality sequence stop: 810.                                   |                        |                |                       |        |                 |
| FEATURES              | Location/Qualifiers  |                        |                |                       |        |                 |
| source                | 1..823   |                        |                |                       |        |                 |
|                       | /organism="Entamoeba histolytica"                                  |                        |                |                       |        |                 |
|                       | /strain="HML:IMSS"   |                        |                |                       |        |                 |
|                       | /db_xref="taxon:5759"  |                        |                |                       |        |                 |
|                       | /clone_lib="Entamoeba histolytica Sheared DNA"                     |                        |                |                       |        |                 |
|                       | /note="Vector: PHOS1; Site_1: Bst I; Constructed at The            |                        |                |                       |        |                 |
|                       | Institute for Genomic Research (ITGR), Rockville, MD.              |                        |                |                       |        |                 |
|                       | Genomic DNA isolated from broth cultures of E. histolytica         |                        |                |                       |        |                 |
|                       | using a method described by Clark and Diamond (Clark,              |                        |                |                       |        |                 |
|                       | C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a            |                        |                |                       |        |                 |
|                       | method for Diamond identification. Exp. Parasitol.                 |                        |                |                       |        |                 |
|                       | 77:450.). The DNA was mechanically sheared to give a               |                        |                |                       |        |                 |
|                       | tight size distribution (~2 kb). The v + i method used for         |                        |                |                       |        |                 |
|                       | the library construction is described in detail in Smith,          |                        |                |                       |        |                 |
|                       | H.O. and Venter, J.C. (Making small insert libraries for           |                        |                |                       |        |                 |
|                       | whole genome shotgun sequencing projects. In Genome                |                        |                |                       |        |                 |
|                       | Sequencing: A Practical Approach, eds. M. Vaubin and B.            |                        |                |                       |        |                 |
|                       | Barrell, Oxford University Press, 1999)."                          |                        |                |                       |        |                 |
| BASE COUNT            | 228 a  |                        |                |                       |        |                 |
| ORIGIN                | 206 c  |                        |                |                       |        |                 |
|                       | 86 g   |                        |                |                       |        |                 |
|                       | 303 t  |                        |                |                       |        |                 |
| Query Match           | 4.2%   | Score 62.2             | DB 17          | Length 823            |        |                 |
| Best Local Similarity | 47.5%  | Pred. No. 8.1e-06      |                |                       |        |                 |
| Matches 184           | Conservative   | 0                      | Mismatches 203 | Indels 0              | Gaps 0 |                 |
| 32                    | AAATGACCCACCTGACGACGACACCGATGATCACCATGACGACCATGATGATGATG           | 91                     |                |                       |        |                 |
| 658                   | ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG             | 599                    |                |                       |        |                 |
| 92                    | AAACAAATGACCTATGCGCCAGTGTGAAGAAAGAACCAACCTCATATGATGATG             | 151                    |                |                       |        |                 |
| 598                   | ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG             | 539                    |                |                       |        |                 |
| 152                   | ACCCACCATGTCATGATGACGACGATGATGATGATGATGATGATGATGATGATG             | 211                    |                |                       |        |                 |
| 538                   | ATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG             | 479                    |                |                       |        |                 |
| 212                   | AACCTTCATCTTGTGCGATTCACACAAAGTGAAGACCATGACGACCATCATG               | 271                    |                |                       |        |                 |
| 478                   | ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG             | 419                    |                |                       |        |                 |

|                       |  |   |  |             |
|-----------------------|--|---|--|-------------|
| OY                    |  | 272   | ATCTGCACATGCGTGGTGATCATGTCCAGCAGGTTGGATTTCATTGGGCAACTGTACAATG  | 331         |
| Dd                    |  | 418   | ATGATGACGATGATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATG | 359         |
| OY                    |  | 332   | CTCACCCAGAAACATCGTGAACCTCGTAGACCCTGCCGACCTCGGTGACCATTAAGSG     | 391         |
| Dd                    |  | 358   | ATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG   | 299         |
| OY                    |  | 392   | GCGTGGTTAATGAGAATTCATCATATG                                    | 418         |
| Dd                    |  | 298   | ATGATGATGACGATGATAATGAAATGATG                                  | 272         |
| RESULT_7              |  | BH139532  | 843 bp DNA linear GSS-07-AUG-2001                              |             |
| LOCUS                 |  | BH139532  |  |             |
| DEFINITION            |  | ENTNM88rTF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.   |  |             |
| ACCESSION             |  | BH139532  |  |             |
| VERSION               |  | BH139532.1 GI:15098593  |  |             |
| KEYWORDS              |  | GSS.  |  |             |
| SOURCE                |  | Entamoeba histolytica.  |  |             |
| ORGANISM              |  | Entamoeba histolytica.  |  |             |
| REFERENCE             |  | Eukaryota; Entamoebidae; Entamoeba.   |  |             |
| AUTHORS               |  | Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.   |  |             |
| TITLE                 |  | determination of clone end sequences from Entamoeba histolytica unpublished (2001)  |  |             |
| JOURNAL               |  | Contact: Brendan J Loftus   |  |             |
| COMMENT               |  | Department of Eukaryotic Genomics<br>The Institute for Genomic Research<br>9712 Medical Center Dr., Rockville, MD 20850, USA<br>Tel.: 301 838 0208<br>Fax: 301 838 3543<br>Email: bjo@loftuselgfr.org<br>Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library<br>Seq primer: M13-Forward<br>Class: Shotgun<br>High quality sequence start: 4<br>High quality sequence stop: 714.<br>Location/Qualifiers<br>1..843<br>organism="Entamoeba histolytica"<br>strain="HMI:IMSS"<br>db_xref="taxon:5759"<br>clone_lib="Entamoeba histolytica Sheared DNA"<br>note=Vector: PHOSI, Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.<br>Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999). |  |             |
| BASE COUNT            |  | 324 a   | 82 c   | 205 g       |
| ORIGIN                |  | 232 t   |  |             |
| Query Match           |  | 4.2%  | Score 62.2,  | DB 17;      |
| Best Local Similarity |  | 47.5%   | Pred. No. 8.3e-06;   | Length 843; |
| Matches 184:          |  | Conservative 0;   | Mismatches 203;  | Indels 0;   |
|                       |  |   | Gaps 0;  |             |
| OY                    |  | 32  | AAGATGACCAACATGACGACCAACGATGATCACCATGACGACCATGATGATGATG        | 91          |
| Dd                    |  | 224   | ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 283         |
| OY                    |  | 92  | AAACAATGACCTATGCCCAGTGTGAATGGAACCAACCCCTCATATGCTAGACGCTTC      | 151         |

Mon Dec 2 12:19:39 2002

us-09-868-760-6.rst

Page 6

[illegible]

RESULT 8  
AV679268/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AV679268 740 bp mRNA linear EST 05-OCT-2000  
AV679268 Nori Satoh unpublished cDNA library Clona intestinalis  
cDNA clone rc1btb5h19 3', mRNA sequence.  
AV679268  
AV679268.1 GI:10117267  
EST.  
Clona intestinalis.  
Clona intestinalis  
Eukaryota; Metazoa; Chordata; urochordata; Ascidiacea; Enterozona;  
Phlebobranchia; Clonidae; Clona.  
1 (bases 1 to 740)  
Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.  
Expressed genes in Clona intestinalis  
Unpublished (2000)  
Contact: Nori Satoh  
Department of Zoology  
Kyoto University  
Kyoto-ku, Kyoto, Kyoto 606-8502, Japan  
Tel.: 81-75-753-4081  
Fax: 81-75-703-1113  
Email: satoh@asciidian.zool.kyoto-u.ac.jp.

| FEATURES   | source                           |
|--|----------------------------------|
| Location/Qualifiers                              | 1. 740                           |
| /organism="Ciona intestinalis"                   |                                  |
| /db_xref="taxon:7719"                            |                                  |
| /clone="rc1b15h19"                               |                                  |
| /clone_lib="Nori Satoh unpublished cDNA library" |                                  |
| /tissue_type="whole animal"                      |                                  |
| /dev_stage="tallbud"                             |                                  |
| /note="vector: pBluescript SK"                   |                                  |
| BASE COUNT                                       | 150 a 216 c 68 g 292 t 14 others |
| ORIGIN   |                                  |

|    | Query Match           | 3.9%   | Score 58.8        | DB 10    | Length 740 |
|----|-----------------------|--|-------------------|----------|------------|
|    | Best Local Similarity | 45.7%  | Pred. No. 6.5e-05 |          |            |
|    | Matches 195           | Conservative 0   | Mismatches 222    | Indels 0 | Gaps 0     |
| QY | 12                    | GTATACGATGGGCAGAACAAAGATGACCAACCATGACGACCAACCGATGATCACCATGA  | 71                |          |            |
|    |                       |  |                   |          |            |
| DB | 550                   | GNGAGATGATGACCTAGATGATGATACCAAGATGATGNNAGATGTGTGCGCAAGTGA    | 491               |          |            |
| QY | 72                    | CGACCATGATGATGATGATGAACAATGCACTATGCCCGAGTGTGAATGGAACCAACC    | 131               |          |            |
|    |                       |  |                   |          |            |
| DB | 490                   | CGAAATATCATACCAACAAATGATGATGACCAAGTATGTATGCCCAAGATGATGA      | 431               |          |            |
| QY | 132                   | TCATATGGCTAGACGACCTTCACCAACCATGTCATGGCAGCATAGATTGTTCACGAAGGG | 191               |          |            |
|    |                       |  |                   |          |            |

|    |     |   |     |
|----|-----|---|-----|
| Db | 430 | CGAAGATGATGACGAAGATGATGACGTGATGATGATGAAGAAGATGATGACGAAGATGA     | 371 |
| Qy | 192 | TCATGGAGCGTTTATCTGAACTTCATCTGTGTGATTCACACAGGAAGAACCATGA         | 251 |
| Db | 370 | TGACGTGATGATGATGACGAAGAAATGATGACGAAGATGATGACGAAGAAAGTATGACGAAGA | 311 |
| Qy | 252 | CGACACCATATGAGACTTCATCTGCATGCTTGTGTGACATGCTCAGCAGGTGTGATTC      | 311 |
| Db | 310 | TGACGAAGATGATGAAGAAGATGATGACGTGATGATGATGACGTGATGATGATGACGAAGA   | 251 |
| Qy | 312 | TATTGGGCAAGCTTACAAATGCTCACCACGAAAAACATGTCACCCCTGGTAGCTCGTGA     | 371 |
| Db | 250 | TGATGACGAAGATGATGACGAAGATGACGAAGATGATGACGAAGATGATGACGTATGATATGC | 191 |
| Qy | 372 | CCGTGTTGACGATGATGAGGGCGTGTGTTAATGAAGTTCATCATTTAGCTTGGTGGACAT    | 431 |
| Db | 190 | TGACGAAGATGATGACGAAGATTAATGACAAAGATGACGACGAAGATGCTCAGCAAGATGA   | 131 |
| Qy | 432 | TGATGCT 438   |     |
| Db | 130 | TGACGAT 124   |     |

[illegible]

```

/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOSt1 Site: 1 Bst I: Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.) The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + l method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1993)."

```

|                           |        |                    |           |             |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match               | 3.9%   | Score 58.4;        | DB 17;    | Length 976; |
| Best Local Similarity     | 45.9%; | Pred. No. 0.00011; |           |             |
| Matches 200; Conservative | 0;     | Mismatches 236;    | Indels 0; | Gaps 0;     |

|    |     |  |     |
|----|-----|--|-----|
| QY | 1   | GATGGGAGCAGTGTAAACGATGGCGAGAAACAAACATGTGCACCCATGTGACGCACACACAGAT | 60  |
| Db | 298 | GAAGATGTAAGAAATGAAAGAACACATGATGTGAAGAGCCGATGTGTGAAGAAAGACGATGAT  | 357 |
| QY | 61  | GATCACCATTAGCACCATTGATGATGATGTGTAACAAATGCATGTGCCCGTGTGAAATG      | 120 |
| Db | 358 | GAAGAAAGCAGTATGAAAGAACACATGTGTGAAGAAAGCCGATGTGTGAAGAAAGACGATGAT  | 417 |
| QY | 121 | GAACCAAAACCTCATGTGGCTAGCAGCCCTGCACCCATGTCATGTGCAGACGATAGAGTTG    | 180 |
| Db | 418 | GAAGCAAAACCTCATGTGGCTAGCAGCCCTGCACCCATGTCATGTGCAGACGATAGAGTTG    | 477 |

[illegible]

| RESULT 10  | LOCUS     | DEFINITION  |
|------------|-----------|---|
| A2551618/c | A2551618  | 843 bp DNA linear GSS 14-NOV-2000   |
|            | ENTDP544R | Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic DNA sequence. |

|           |   |             |
|-----------|---|-------------|
| VERSION   | A2551618.1  | GT:11176919 |
| KEYWORDS  | GSS   |             |
| SOURCE    | Entamoeba histolytica.  |             |
| ORGANISM  | Entamoeba histolytica.  |             |
| REFERENCE | Eukaryota; Entamoebidae; Entamoeba.                             |             |
| AUTHORS   | 1 (bases 1 to 843)  |             |
| TITLE     | Loftus,B., Van Aken,S. and Fraser,C.                            |             |
| JOURNAL   | Determination of clone end sequences from Entamoeba histolytica |             |
| COMMENT   | HMI:IMSS sheared DNA library                                    |             |
|           | unpublished (2000)  |             |
|           | Contact: Brendan J Loftus                                       |             |

The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: [bjloftus@tigr.org](mailto:bjloftus@tigr.org)  
Clones are derived from the *Entamoeba histolytica* HM1:IMSS sheared  
DNA library  
Class: Shotgun  
Seq primer: M13-Reverse  
High quality sequence start: 39  
High quality sequence stop: 878.  
Location/Qualifiers  
1..843

```
/organism="Entamoeba histolytica"  
/strain="HM1:IMSS"  
/db_xref="taxon:5759"  
/clone_lib="Entamoeba histolytica Sheared DNA"
```

/note=Vector: Site\_1: Bat I: Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp. Parasitol.* 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. in *Genome Sequencing: A Practical Approach*, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).

|                       |                 |                   |           |             |
|-----------------------|-----------------|-------------------|-----------|-------------|
| Query Match           | 3.88;           | Score 57.2;       | DB 17;    | Length 843; |
| Best Local Similarity | 43.28;          | Pred. No. 0.0002; |           |             |
| Matches 324;          | Conservative 0; | Mismatches 423;   | Indels 3; | Gaps 1;     |

|    |     |  |     |
|----|-----|--|-----|
| QY | 188 | AGGGTCATGAGCTATTTATATCAGAACTTCCATCTTTCGGATTCAACACAAAGAAAGACC   | 247 |
| Db | 777 | AAGATGATGAAGAAGTGAATGAGAAAGAAAGAGTGTGATGAAGATGAAAGATGAAGAGC    | 718 |
| QY | 248 | ATGACGACCAACCATGATGAGCTCTCATCGACATGCTGGTGACATGTGACAGAGTTTGG    | 307 |
| Db | 717 | ATGAAACACACGAAATGTGATTAG- --AGATGATGATGATGATGAAGAAAGAAAGATGATG | 661 |
| QY | 308 | ATTCTATTGCGAACTGTCATATGCTCACCCAGAAAAAATGCTGACCCCTGGGACCTCG     | 367 |
| Db | 660 | ATGAAGATGATGAGAACGCTGAAAGACGACGAAATGTGAATTGAAAGATGATGATGATGATG | 601 |
| QY | 368 | GTCACCTGGTTCAGCTGTATAGGCGCGGTGTTAATGAAAGTTCACTATTPTTGGCTGGTGG  | 427 |
| Db | 600 | AAGAAGAGATGATGATGAGAAAGAAAGAAAGATGATGATGAAGATGATGATGAAGAAAG    | 541 |
| QY | 428 | ACATGTATGTTCACAGACCAAAACCCGAAAGCTGCATATGGACCTCAATGACATATTTC    | 487 |
| Db | 540 | ATGAAGAAATGATGAAAGAAAGTATGATGAAGATGATGATGATGAAGAAAGAAAGATGATG  | 481 |
| QY | 488 | AAGGAGTCACACCATGCTGTATCCCGACGACTAATATGCTGTGTATTATGTC           | 547 |
| Db | 480 | AAGATGTGAAGACGCTGAAAGCACGCAATTTGAATTTGAAGATGATGATGAAGAAAG      | 421 |
| QY | 548 | ATGGAAGAACTCGCCACAGAAACGACGCTGTACATCACAGCTTAGAGAAAGATAAA       | 607 |
| Db | 420 | AAGATGTGTGTGAAGAGAAAGAAAGTGTGATGATGAAGATGATGAAGAAAGAGTGAAG     | 361 |
| QY | 608 | CTGAGCATTAATGCCCCATTGTGACGTAGATCTAATPACACACCAACCAAGCTTTTCATC   | 667 |
| Db | 360 | ATATGATGTGAAGAGAGATGATGAAGTGTGATGATGAAGAAAGATGATGAAGAAAGAG     | 301 |
| QY | 668 | ATCATGTCCACGAAACCATCGATTTCAAPACAGTTGGTTATGTGTGACCTTGAAGTCTCT   | 727 |
| Db | 300 | ATGAAGAGAAAGAAAGTGTGATGAAGAGTGTGATGAAGAAAGAAAGATGATGAAGAGC     | 241 |
| QY | 728 | ACCATTTAAGAGGATTTAATGTAAAGTGAACCAAAAGATCATCTTCATGACCTACAGA     | 787 |
| Db | 240 | ACGAATTTGATTTAATGAAGTGTGATGAAGAAAGTGTGAAAGAAAGATGATGAAGAAAGATG | 181 |
| QY | 788 | TCTAGCCCAACGGTGACCTGACCAAGTGAATTAACCTCGGTGCTAAATATGATCTTC      | 847 |
| Db | 180 | ATGAAGAAAGAAAGTGTGATGATGATGATGATGAAGAAAGATGATGATGAAGAGATGAAG   | 121 |
| QY | 848 | ATGAAGATTTACCACTAGTAGTGGGTGATCTTAGAGAGATATTCCAGATGATGACCATGGC  | 907 |
| Db | 120 | AAGATGAAAGTGTGATGAAGAAAGTGAAGATGATGATGAAGATGATGATGAAGATGATG    | 61  |
| QY | 908 | TTGTCAATGAAGCCACAGATATTCCTTGA                                  | 937 |
| Db | 60  | AAGATGATGATGAAGACGCAATTTAATTGA                                 | 31  |

|            |   |
|------------|---|
| RESULT 11  |   |
| LOCUS      | B0731479  |
| DEFINITION | B0731479 887 bp mRNA linear EST 16-JUL-2002<br>AGENCOURT_8216778 NICHD XSC Embd Xenopus laevis cDNA clone   |
| ACCESSION  | IMAGE:4930330 5', mRNA sequence.  |
| VERSION    | B0731479  |
| KEYWORDS   | B0731479.1 GI:21870376  |
| SOURCE     | EST.  |
| ORGANISM   | African clawed frog.<br>Xenopus laevis<br>Xenopus laevis<br>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;<br>Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;<br>Xenopodinae; Xenopus.<br>1 (bases 1 to 887)<br>NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .<br>National Cancer Institute, Cancer Genome Anatomy Project (CGAP),<br>Tumor Gene Index<br>Unpublished (1997) |
| JOURNAL    | Contact: Robert Strusberg, Ph.D.<br><a href="mailto:rs@lsc.ucsf.edu">rs@lsc.ucsf.edu</a>  |
| COMMENT    |   |

CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
plate: LLAM10852 row: 1 column: 11  
high quality sequence stop: 603.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source   | 1. .887             |

```

/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4930330"
/clone_lib="NICHD XCC Embd4"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: whole embryo; Vector: PCMV-SPORT6; Site:1;
Note: Site:2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus gene collection (XGC
) library."

```

|    | Query Match  | Best Local Similarity   | Score 57.2; 46.9%; | Pred. No. 0.0021; | Length 887; |
|----|--------------|---|--------------------|-------------------|-------------|
|    | Matches 179; | Conservative 0;   | Mismatches 203;    | Indels 0;         | Gaps 0;     |
| OY | 32           | AAGTAGCCACCACTGACGACCGACCACGATGATACCATGACGACCATGATGATGATG         | 91                 |                   |             |
| Dd | 350          | ACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 409                |                   |             |
| OY | 92           | AAACAATGCACACTTGTGCCCACTGTGGAAATGGAACCAAACCCCTAATTGGCTAGCAGCCCTTC | 151                |                   |             |
| Dd | 410          | ATGATGACGACGATGATGATGATGACGACGATGATGATGATGATGATGATGATGATGATG      | 469                |                   |             |
| OY | 152          | ACCAACCACTGTCACATGACGATGATGATTTCTCACAAAAGGTCATGACGACCTTTATCTAC    | 211                |                   |             |
| Dd | 470          | ACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 529                |                   |             |
| OY | 212          | AACTTCATCTTTGCGGATTCACACAACTGAAGAACCATGACGACCACTCATGAGACTTC       | 271                |                   |             |
| Dd | 530          | ACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG         | 589                |                   |             |
| OY | 272          | ATTCGCACATGCTTGTTGGTACATGTCACGACGGTTGTGATTTCTATTGGCGAACGTGTACATG  | 331                |                   |             |
| Dd | 590          | TACAAAAAGATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 649                |                   |             |
| OY | 332          | CYCACCCAGAAAACATGCTGACCCCTGTGTGACCTCGTGACCTGGTTGACGATGATGAGG      | 391                |                   |             |
| Dd | 650          | ATGACGACCATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 709                |                   |             |

|    |     |    |     |     |    |    |    |    |    |     |
|----|-----|----|-----|-----|----|----|----|----|----|-----|
| Qy | 392 | GC | GT  | GT  | TA | TG | AA | GT | CA | 413 |
|    |     |    |     |     |    |    |    |    |    |     |
| Db | 710 | AT | GAC | CAG | AA | GA | TG | AT | GA | 731 |

| RESULT 12 | LOCUS                                     | DEFINITION   |
|-----------|---|--|
| Bj374097  | 825 bp                                    | linear EST 08-MAR-2002   |
| Bj374097  | Dicystostellum discoidum cDNA library, CF | Dicystostellum discoidum cDNA clone ddcdbis 3', mRNA sequence. |

|           |  |
|-----------|--|
| ACCESSION | BJ374097   |
| VERSION   | BJ374097.1   |
| KEYWORDS  | GI:19283480  |
| SOURCE    | EST  |
| ORGANISM  | Dictyostellium discoideum.                                       |
| REFERENCE | Dictyostellium discoideum.                                       |
| AUTHORS   | Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.              |
| TITLE     | 1 (bases 1 to 825)   |
|           | Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-I, T.            |
|           | Full length cDNA of Dictyostellium discoideum at the culmination |
|           | stage  |

unpublished (2002)  
Contact: Tadasu Shin-i  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima,  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp

```

FEATURES
source
1. . 825
Location/Qualifiers
format= "D:\structure\

```

```

/organism="Dictyostellum discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc6b15"
/clone_11b="Dictyostellum discoideum cDNA library, CF"
/sex="mat A"
/dev_stage="Culmination stage"
BASE COUNT      190 a      113 c      133 g      229 t      160 others
ORIGIN

```

| Query Match | Similarity   | 3.8%  | Score 56.8       | DB 13    | Length 825 |
|-------------|--------------|---|------------------|----------|------------|
| Best Local  | Similarity   | 36.1%   | Pred. No. 0.0026 |          |            |
| Matches 255 | Conservative | 0   | Mismatches 451   | Indels 7 | Gaps 2     |
| OY          | 203          | TTTATCTAGACTTTCATCTTGTGCGATTCACACACAGTGAAGACCATGACGACCACCATC  | 262              |          |            |
| Db          | 96           | TTTGAACTGAACACGCTTGAGCTAGAACCTCTTGAGCTTGAACCTTGAGACTGCTTGAG   | 155              |          |            |
| OY          | 263          | ATGACATCTCATCTGCACATGCTTGGTGCATCTGCAGAGCTTGATCTATTGGCGAAC     | 322              |          |            |
| Db          | 156          | TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 215              |          |            |
| OY          | 323          | TGTTCATATGTCACCCGAAAAACATGCTGCGACCTGGTGCAGCTGCGACTGGTTGAGC    | 382              |          |            |
| Db          | 216          | ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 275              |          |            |
| OY          | 383          | ATGATGAGGCGCTGGTTAAATGAAGTTTCATCTATTATGCTTGGTGGACATTGATGCT    | 440              |          |            |
| Db          | 276          | ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG     | 335              |          |            |
| OY          | 441          | AGCACCAACACGAGACTCTCATTTGGACACTCATCTATTTTACAGAGGATGCACAC      | 500              |          |            |
| Db          | 336          | NGCNCNTGANNNGNNGCNCNNNNNCNTGAGAGNACNNCNCNCNCNCNCNTNNTWANN     | 395              |          |            |
| OY          | 501          | CGATGCTGATACCCCGACGAGTAAATGCCCTGGTGTGTATTGGTCATGAAAGAGCTG     | 560              |          |            |
| Db          | 396          | ANCGACNNCNCNNNAGNANCNTNANCNCCTGCTCNCNCNCNNAGNANNNTNANCNCNCTG  | 455              |          |            |
| OY          | 561          | CCGAGAAAACGACGACTGCTCTACATCAGCAGCTAGAGAGAATATAAACCTGAGCTTATG  | 620              |          |            |
| Db          | 456          | NTNATCANCNCG-----NNCNCNCNTGAGNANCNTNANCAACTCNCNNNNCNCNAGNANCN | 510              |          |            |
| OY          | 621          | CCATTGAGAGTAAAGTCTAATACACACCAACCAAAGGCTCTTCATCATCATGTCACGG    | 680              |          |            |

|    |     |  |     |
|----|-----|--|-----|
| Db | 511 | TCNTCNTNANCAACNNCNCNACNNCTCNCNTCNCNTNANNANCAANCNNCNCNT           | 570 |
| Oy | 681 | AACCATGCATTTGCAACAGAGTTGGTTATGGACCTTGAAGTGCTCCATCCATTTAGAGGG     | 740 |
| Db | 571 | NCNTCNTNTNATGATGAACAACATCNCNCTTCCTTNATNATGATGANNATCATGATGATGATNA | 630 |
| Oy | 741 | ATTTTAATTAAGTGATGACCAACAAGATCATCTCATGACGTACAGATCTACGCCAAGG       | 800 |
| Db | 631 | NCNTCNTCTTATGATGATGATGATNATNANCNTCNCNTNATCANCNTCNCNTNTNTGATNA    | 690 |
| Oy | 801 | TGACCTGCACAGTGGATGATTAACCTGGGGCTTAAATFATGATCCATCGAAGATTTACCA     | 860 |
| Db | 691 | TNATCTTCNTCNCNTCTTNATNATNATCATCATCANCNTCNTNATNATGATGATGATCA      | 750 |
| Oy | 861 | CAGTGAGTGGGTACTGAGAGATATTCACAGATGAGACATGGCGTTCATCATC             | 917 |
| Db | 751 | TGATCNTCNTNATGATGATCATCANCCTTCNCNTTNATGATGATGATGATGATGATGA       | 807 |

| RESULT 13  |             |                    |                |        |                 |
|------------|-------------|--------------------|----------------|--------|-----------------|
| BO834753/C | BO834753    | 644 bp             | mRNA           | linear | EST 08-AUG-2002 |
| LOCUS      | PO.ad.03M07 | TEXEL              | Psoroptes ovis | mixed  | Psoroptes ovis  |
| DEFINITION | PO.ad.03M07 | 5', mRNA sequence. |                |        | CDNA clone      |
| ACCESSION  | BO834753    |                    |                |        |                 |
| VERSION    | BO834753.1  | GI:22139067        |                |        |                 |
| KEYWORDS   | EST.        |                    |                |        |                 |
| SOURCE     | sheep       | scab mite.         |                |        |                 |

|           |  |
|-----------|--|
| REFERENCE | 1 (bases 1 to 644)   |
| AUTHORS   | Keynon, F., Welsh, M., Parkinson, J., Whitton, C., Blaxter, M. and Knox D. P.  |
| TITLE     | Expressed sequence tag survey of gene expression in the scab mite <i>Psoroptes ovis</i> - allergens, proteinases and free radical scavengers |
| JOURNAL   | Unpublished (2002)   |
| COMMENT   | Contact: Blaxter ML  |

Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JF, UK.  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
The library was prepared for Dr David Knox, Moredun Research  
Institute, UK Sequencing was performed by Mark Welsh, ICAPB,  
Edinburgh  
PCR Primers  
FORWARD: M13R  
BACKWARD: T7PL  
Plate: 03 row: A column: 07  
High quality sequence stop: 536.

```

FEATURES
source
    location/Qualifiers
        1. 644
            /organism="Psoroptes ovis"
            /db_xref="taxon:83912"
            /clone_1b="Po.ad.03A07"
            /clone_1lb="Psoroptes ovis mixed"
            /sex="Mixed"
            /dev_stage="Mixed (predominantly adult)"
            /note="Vector: lambda Triplex2, Site_1: SfilA, Site_2:
            SfilB, Psoroptes ovis is the infective agent causing sheep
            scab, a serious dermatitis in sheep. The cDNA was
            obtained from mites harvested from the skin of previously
            unexposed sheep harbouring a moderate challenge
            infection. The cDNA was cloned using the Clontech lambda
            Triplex2 Smart cDNA cloning system."
205 a 133 c 123 g 183 t

```

|                           |        |                    |           |             |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match               | 3.88;  | Score 56.6;        | DB 14;    | Length 644; |
| Best Local Similarity     | 47.68; | Pred. No. 0.00024; |           |             |
| Matches 167; Conservative | 0;     | Mismatches 184;    | Indels 0; | Gaps 0;     |

[illegible]

|            |   |
|------------|---|
| RESULT     | 14  |
| B0834718/c |   |
| LOCUS      |   |
| DEFINITION | B0834718 661 bp mRNA linear EST 08-AUG-2001   |
| ACCESSION  | Po.ad_02F05_TEXF1 Psoroptes ovis mixed Psoroptes ovis cDNA clone Po.ad_02F05 5', mRNA sequence. |
| VERSION    | B0834718  |
| KEYWORDS   | B0834718.1 GI:22139032<br>EST.  |
| SOURCE     | sheep scab mite.<br><i>Psoroptes ovis</i>   |
| ORGANISM   |   |

|           |  |
|-----------|--|
| REFERENCE | 1 (bases 1 to 661)   |
| AUTHORS   | Kenyon, F., Welsh, M., Parkinson, J., Whilton, C., Blaxter, M. and Knox, D. P. |
| TITLE     | Expressed sequence tag survey of gene expression in the scab mite              |
| JOURNAL   | Psoralea ovis - allergens, proteinases and free radical scavengers             |
| COMMENT   | Unpublished (2002)<br>Contact: Blaxter ML                                      |

Institute of Cell, Animal and Population Biology  
University of Edinburgh  
Asmuthw Labs, King's Buildings, West Mains Road, Edinburgh, EH9  
3JF, UK.  
Tel: +44 131 650 6760  
Fax: +44 131 670 5450  
Email: mark.blaxter@ed.ac.uk  
The library was prepared for Dr David Knox, Moredun Research  
Institute, UK Sequencing was performed by Mark Welsh, IACBP,  
Edinburgh.  
PCR primers  
FORWARD: M13R  
BACKWARD: T7PL  
Plate: 02, row: F column: 05  
High quality sequence stop: \$33.  
Location/Qualifiers  
1. .661

```

/organism="Psoroptes ovis"
/db_xref="taxon:83912"
/clone="Po_ad_02F05"
/clone.lib="Psoroptes ovis mixed"
/sex="Mixed"
/dev_stage="Mixed (predominantly adult)"
/note="Vector: Lambda triplex; Site_1: Sf1a; Site_2:

```

slf1b; psoroptes ovis is the infective agent causing sheep scab, a serious dermatitis in sheep. The cDNA was obtained from mites harvested from the skin of previously unexposed sheep harbouring a moderate challenge infection. The cDNA was cloned using the Clontech Lambda Triplex2 Smart cDNA cloning system."

BASE COUNT 211 a 136 c 130 g 184 t

Query Match 3.8%; Score 56.6; DB 14; Length 661;  
Best Local Similarity 47.6%; Pred. No. 0.00024;  
Matches 167; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

OY 607 ACTGAGCATTTCGCCCTGTGACGTAAGATCTATACACCAACCAAGCCTTCAT 666  
DB 505 ACCAAGAGCACTGATTGTGTCATCGACATGATGATGATGATGATGATGATG 446  
OY 667 CATCATGTCACGACGACCATGATTTCAACCAAGTGTGTGAGCCTTGAAGTCC 726  
DB 445 CACCATGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 386  
OY 727 TACCATTTAGAGGAGTTTAATGTAAGTATGACCAACCAAGTATCTCCATGAC 786  
DB 385 CAAAGTGTATGACCGACGATGATGATGATGATGATGATGATGATGATGAT 326  
OY 787 ATCTAGCCCAAGCGTACGATGATGATGATGATGATGATGATGATGATGAT 846  
DB 325 GAGGACGTCACGATGATGATGATGATGATGATGATGATGATGATGATGAT 266  
OY 847 CATGAAGATTACACAGTATGATGATGATGATGATGATGATGATGATGATG 906  
DB 265 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 206  
OY 907 GTGTCAATGAAGACGACATATCTCGATGATGATGATGATGATGATGATGAT 957  
DB 205 GATGACGACGAAATGACCATGCTGATGATGATGATGATGATGATGATGAT 155

RESULT 15  
BM161314 677 bp mRNA linear EST 04-DEC-2001  
LOCUS EST63837 PyBS Plasmodium yoelli yoelli cDNA clone PYCKD34 5' end,  
DEFINITION mRNA sequence.  
ACCESSION BM161314.1 GI:17306995  
VERSION EST  
KEYWORDS Plasmodium yoelli yoelli.  
SOURCE Plasmodium yoelli yoelli.  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 677)  
AUTHORS Carlson,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,  
Fraser,C.W. and Carucci,D.J.  
JOURNAL Plasmodium yoelli EST Project at TIGR  
COMMENT Unpublished (2001)  
Contact: Jane Carlson  
Parasite Genomics Group  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-530-9319  
Fax: 301-838-0208  
Email: carlton@tigr.org  
For clone info, please contact the Malaria Research and Reference  
Reagent Resource Center, ATCC  
http://www.malaria.mr4.org/mr4pages/index.html  
Seq primer: ADF.

FEATURES  
Source Location/Qualifiers  
1..677  
/organism="Plasmodium yoelli yoelli"  
/strain="17XL"  
/db\_xref="taxon:73239"  
/clone="PYCKD34"  
/clone\_lib="PyBS"  
/dev\_stage="Asexual blood stages"

/lab host="E. coli XL-1 Blue"  
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven leamini were treated with Pfu DNA polymerase and EcoRI adapters ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HydrilAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HydrilAP vector and plasmid DNA isolated."

BASE COUNT 288 a 50 c 186 g 153 t

Query Match 3.8%; Score 56.4; DB 13; Length 677;  
Best Local Similarity 48.4%; Pred. No. 0.00028;  
Matches 156; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

OY 598 GAAGATAAATGAGCATTTATGCCATTTGTGACGTAAGATCTAATACACCAACCAAG 657  
DB 91 GAAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 150  
OY 658 GCCTTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717  
DB 151 ACAAGGAAATGAGCAGAGAGAGAGAGATGATGATGATGATGATGATGAT 210  
OY 718 GAAGTCTCTACCATTTAGAGGATTTAATGTAAGTATGATGATGATGATGAT 777  
DB 211 GAGCAGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 270  
OY 778 GAGGTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
DB 271 GAGCATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 330  
OY 838 TATGATCTCAAGATTAATACACGATGATGATGATGATGATGATGATGATGAT 897  
DB 331 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 390  
OY 898 GACCATGCGCTGTCAATGA 919  
DB 391 GACGATGAAGATTTGAAGATA 412

Search completed: November 29, 2002, 10:02:51  
Job time : 2074 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 07:43:13 ; Search time 2829 Seconds

(without alignments)  
15338.382 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gaggggagcagtgtaacga.....gctgtgtatagagcagcgca 1491

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank :  
1: gb\_ba :  
2: gb\_hlg :  
3: gb\_in :  
4: gb\_ov :  
5: gb\_ov :  
6: gb\_ov :  
7: gb\_ov :  
8: gb\_ov :  
9: gb\_ov :  
10: gb\_ov :  
11: gb\_ov :  
12: gb\_ov :  
13: gb\_ov :  
14: gb\_ov :  
15: gb\_ov :  
16: gb\_ov :  
17: gb\_ov :  
18: gb\_ov :  
19: gb\_ov :  
20: gb\_ov :  
21: gb\_ov :  
22: gb\_ov :  
23: gb\_ov :  
24: gb\_ov :  
25: gb\_ov :  
26: gb\_ov :  
27: gb\_ov :  
28: gb\_ov :  
29: gb\_ov :  
30: gb\_ov :  
31: gb\_ov :  
32: gb\_ov :  
33: gb\_ov :  
34: gb\_ov :  
35: gb\_ov :  
36: gb\_ov :  
37: gb\_ov :  
38: gb\_ov :  
39: gb\_ov :  
40: gb\_ov :  
41: gb\_ov :  
Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1          | 1484.2 | 99.5        | 1700   | 3     | AF273766    |
| 2          | 88.4   | 5.9         | 115758 | 9     | AC104634    |
| 3          | 74.8   | 5.0         | 186278 | 9     | AC079176    |
| 4          | 69.6   | 4.7         | 272545 | 2     | AC090533    |
| 5          | 66.4   | 4.5         | 143716 | 2     | AL627248    |
| 6          | 65.2   | 4.4         | 222546 | 2     | AC096302    |
| 7          | 64     | 4.3         | 272545 | 2     | AC090533    |
| 8          | 63.6   | 4.3         | 163243 | 9     | AL513304    |
| 9          | 59.6   | 4.0         | 211001 | 2     | AC018695    |
| 10         | 58.4   | 3.9         | 61052  | 2     | AC117074    |
| 11         | 58.4   | 3.9         | 61052  | 2     | AC123513    |
| 12         | 58.4   | 3.9         | 163690 | 2     | AC131509    |
| 13         | 57.2   | 3.8         | 51939  | 9     | AC099341    |
| 14         | 56.8   | 3.8         | 176480 | 2     | AC104805    |
| 15         | 56.6   | 3.8         | 1592   | 8     | PAB296740   |
| 16         | 55.6   | 3.7         | 160888 | 9     | AF046872    |
| 17         | 55.6   | 3.7         | 160888 | 9     | AC025540    |
| 18         | 54.8   | 3.7         | 163601 | 2     | AC079031    |
| 19         | 54.6   | 3.7         | 163601 | 2     | AC025224    |
| 20         | 54.6   | 3.7         | 163601 | 2     | AC096530    |
| 21         | 54.4   | 3.6         | 172853 | 9     | AC084361    |
| 22         | 54.2   | 3.6         | 205429 | 2     | AC026358    |
| 23         | 54     | 3.6         | 205429 | 2     | AC005506    |
| 24         | 53.8   | 3.6         | 162428 | 9     | AC096326    |
| 25         | 53.8   | 3.6         | 171260 | 2     | AC026743    |
| 26         | 53.6   | 3.6         | 49144  | 2     | AC131380    |
| 27         | 53.4   | 3.6         | 164399 | 3     | PFMAL3P6    |
| 28         | 53     | 3.6         | 59762  | 8     | AB023032    |
| 29         | 52.8   | 3.5         | 91048  | 9     | AC008619    |
| 30         | 52.6   | 3.5         | 114771 | 9     | HS569019    |
| 31         | 52.6   | 3.5         | 157420 | 2     | AL845482    |
| 32         | 52.6   | 3.5         | 172387 | 2     | AC104753    |
| 33         | 52.6   | 3.5         | 172387 | 2     | AC018875    |
| 34         | 52.6   | 3.5         | 188121 | 9     | AC012451    |
| 35         | 52.6   | 3.5         | 218815 | 2     | AL807384    |
| 36         | 52.4   | 3.5         | 306    | 8     | AF231791    |
| 37         | 52.4   | 3.5         | 550    | 11    | G67741      |
| 38         | 52.4   | 3.5         | 60953  | 2     | AC100470    |
| 39         | 52.4   | 3.5         | 142447 | 10    | AL627204    |
| 40         | 52.4   | 3.5         | 254197 | 2     | AF389853    |
| 41         | 52.2   | 3.5         | 115489 | 2     | AC117072    |
| 42         | 52.2   | 3.5         | 186562 | 2     | AL451079    |
| 43         | 52.2   | 3.5         | 200368 | 9     | AC096541    |
| 44         | 52     | 3.5         | 83122  | 8     | ATT20010    |
| 45         | 52     | 3.5         | 145524 | 2     | AC016777    |

#### ALIGNMENTS

RESULT 1  
LOCUS AF273766 1700 bp mRNA linear INV 20-MAR-2001  
DEFINITION Perna canaliculus perna precursor, mRNA, complete cds.  
ACCESSION AF273766  
VERSION AF273766.1 GI:13383377  
KEYWORDS  
SOURCE Perna canaliculus.  
ORGANISM Perna canaliculus.  
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
Mytiloidea; Mytilidae; Perna.  
REFERENCE  
1 (bases 1 to 1700)  
AUTHORS Scotti,P.D., Dearing,S.C., Greenwood,D.R. and Newcomb,R.D.  
TITLE Perna: a novel, self-aggregating haemolymph protein from the New Zealand green-lipped mussel, Perna canaliculus (Bivalvia);

Myllilä, J. Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 128 (4), 767-779 (2001).

MEDLINE 21186417  
 PUBMED 11290459  
 2 (bases 1 to 1700)

Scott, P. D., Deating, S. C., Greenwood, D. R. and Newcomb, R. D. Direct Submission  
 Submitted (31-MAY-2000) The Horticulture and Food Research Institute of New Zealand Ltd, 120 Mt. Albert Road, Auckland, New Zealand

| FEATURES | Location/Qualifiers           |
|----------|-------------------------------|
| source   | 1..1700                       |
|          | /organism="Perna canaliculus" |
|          | /db_xref="taxon:38949"        |
| CDS      | 34..11587                     |

```

/note="haemolymph protein; N-terminus determined by
microsequencing of perlin; DEQCNDDCN and HPIC-purified
CNBR and tryptic digest fragments: ASSLNHHVHG; VVNEVHH;
GNSHREIVH; YHGCHDA; QGSHVEVSERVACVYGRA"
/codon_start=1
/evidence=experimental
/product="perlin precursor"
/protein_id="AAK0952.1"
/db_xref="GI:13383378"
/translation="MKLLLSLVFAALALQVRADQEQNDQNKDHHDDHDDHDD
DHHDDDEMHVAYQCEMERPNHMASSLNHHVHGSITELSQGSAVYLEHLTGFTSTSD
HDDHNGGILHMLGSDMSAGCSDSTGELYNHAKPKHNDPGDLDYDDDDGVVNEVHNEA
WLDIDGTPRNTGALIGHSMTITLQGSHTDTPASTRAACCVIGSHKARETVAALHDEL
BEDKTEGADGVRSNTHOKALAHNHVGTITDEKQGVXDLEVSYHLKGFVSDNDHDL
HLHDVQIYANGDLTSCGCDNLGKATVREHVEVHSLDLDHDDHGGVSHESHRSWIN
LHGDSDYSGSLAIHQSDHNLGKSAKTAACCVIAGSGNSHREIVHACVYARSTEGGSI
HNVASSTFEEDTBCGSGTHMTADLKGNSVSPDSLHHNHGQVLEMGSDNGHSGTGRAY
HNDHDAHDFKRFGLDGLVDYIDDSHGIVHSTRTDHLNVEDLNKRSLYIQGSHVEVSER
VACVYGRA"
/sig_peptide
/mat_peptide
34. . 93
94. .1384
/product="perlin"
1650. .1653
poly_a_signal
BASE COUNT 515 a 375 G 379 g 431 t
ORIGIN

```

[illegible]

|    |      |  |      |
|----|------|--|------|
| QY | 421  | TGGTTGGACATTTGATGGTACAGACACCCAAACACCGAAGCTCTATTTGGACACTCAATACACT | 480  |
| Db | 514  | TGGTTGGACATTTGATGGTACAGACACCCAAACACCGAAGCTCTATTTGGACACTCAATACACT | 573  |
| QY | 481  | ATTTTCAAGGGAGTCAACCGATGCTGATACCCACGACAGTAGAATGCGCTGTTGTGT        | 540  |
| Db | 574  | ATTTTCAAGGGAGTCAACCGATGCTGATACCCACGACAGTAGAATGCGCTGTTGTGT        | 633  |
| QY | 541  | ATTGGTCATGAAAAAGCTGCGCCGAAACAGACAGCTGCTCTACATCACAGACTAGAGAA      | 600  |
| Db | 634  | ATTGGTCATGAAAAAGCTGCGCCGAAACAGACAGCTGCTCTACATCACAGACTAGAGAA      | 693  |
| QY | 601  | GATPAAAACAGACATTTATGCCATTTGGAGCTAAAGATCTTAATACACCAACAAAGGCT      | 660  |
| Db | 694  | GATPAAAACAGACATTTATGCCATTTGGAGCTAAAGATCTTAATACACCAACAAAGGCT      | 753  |
| QY | 661  | CTTTCATCATCATGCTCCACGGAACCATGATTTCAACCAAGTTGGTTATGGACCTTGA       | 720  |
| Db | 754  | CTTTCATCATCATGCTCCACGGAACCATGATTTCAACCAAGTTGGTTATGGACCTTGA       | 813  |
| QY | 721  | GTCGCTACCATTTAGAGGGATTTTAATGTAAGTAGTAGACCAACAAAGATCATCTCCATGAC   | 780  |
| Db | 814  | GTCGCTACCATTTAGAGGGATTTTAATGTAAGTAGTAGACCAACAAAGATCATCTCCATGAC   | 873  |
| QY | 781  | GTCACAGATCTAGCCCAAGGGTAGACCTGACCAAGTAGTAGATTAACCTCGTGCTTAATAT    | 840  |
| Db | 874  | GTCACAGATCTAGCCCAAGGGTAGACCTGACCAAGTAGTAGATTAACCTCGTGCTTAATAT    | 933  |
| QY | 841  | GATCCATCATGAAGATTATACCAAGTAGTAGTGGGGTAGCTAGAGATATTACAGATGATGAC   | 900  |
| Db | 934  | GATCCATCATGAAGATTATACCAAGTAGTAGTGGGGTAGCTAGAGATATTACAGATGATGAC   | 993  |
| QY | 901  | CATGGCGTTGTCAATGAAAAGCCACAGATATTCTCGATCAATATCTCGGTGATGACAGT      | 960  |
| Db | 994  | CATGGCGTTGTCAATGAAAAGCCACAGATATTCTCGATCAATATCTCGGTGATGACAGT      | 1050 |
| QY | 961  | GTCCTGGAGCGTTTATTGCCCTTACCACAAAGAGACATCTTCATPAAAAGTCCAAAATT      | 1020 |
| Db | 1054 | GTCCTGGAGCGTTTATTGCCCTTACCACAAAGAGACATCTTCATPAAAAGTCCAAAATT      | 1111 |
| QY | 1021 | GCGCTGTGTGTCTATAGAGCGTGGACAGAGCCATCCAAATTTTTCACAGACTPAAATGT      | 1080 |
| Db | 1114 | GCGCTGTGTGTCTATAGAGCGTGGACAGAGCCATCCAAATTTTTCACAGACTPAAATGT      | 1187 |
| QY | 1081 | GTTTGTCAAGCTTAATACGAATCTACGTGGTTTATACACCATGCTCTGTTCTTAATACA      | 1140 |
| Db | 1174 | GTTTGTCAAGCTTAATACGAATCTACGTGGTTTATACACCATGCTCTGTTCTTAATACA      | 1220 |
| QY | 1141 | TTTGACACAGACCCCTGGAGGATCAACACATATGACGGCTATCTCAAAAGATTAAAGCTT     | 1200 |
| Db | 1234 | TTTGACACAGACCCCTGGAGGATCAACACATATGACGGCTATCTCAAAAGATTAAAGCTT     | 1280 |
| QY | 1201 | AGTAGAGACTTGTACATCATGCTGATGGTGTGACGCTCCATGAATGGGAGATATGTCC       | 1260 |
| Db | 1294 | AGTAGAGACTTGTACATCATGCTGATGGTGTGACGCTCCATGAATGGGAGATATGTCC       | 1330 |
| QY | 1261 | CATGGCTGTACCTTAAAGCAAGATTGACCATGGTATGATGATGCTCATGACCCCAA         | 1330 |
| Db | 1354 | CATGGCTGTACCTTAAAGCAAGATTGACCATGGTATGATGATGCTCATGACCCCAA         | 1410 |
| QY | 1321 | AGACCTGTGACCTTGGTATGTATTAAATCATTTCCCATGGATCTTATTTCAACTAGA        | 1380 |
| Db | 1414 | AGACCTGTGACCTTGGTATGTATTAAATCATTTCCCATGGATCTTATTTCAACTAGA        | 1470 |
| QY | 1381 | ACCTTTATCATCTTAAATGTAAGATCTTAAAGCAGCTTCCCTGATGATTATGACGGGC       | 1440 |
| Db | 1474 | ACCTTTATCATCTTAAATGTAAGATCTTAAAGCAGCTTCCCTGATGATTATGACGGGC       | 1510 |
| QY | 1441 | GGACATAGGTCAGAGTAGAGAGGTTGCTGTGTCTTAATAGAGGGGCA                  | 1491 |
| Db | 1534 | GGACATAGGTCAGAGTAGAGAGGTTGCTGTGTCTTAATAGAGGGGCA                  | 1584 |

|            |   |             |     |        |                 |
|------------|---|-------------|-----|--------|-----------------|
| RESULT 2   |   |             |     |        | PRI 10-AUG-2002 |
| ACI04634/c | ACI04634  | 115758 bp   | DNA | linear |                 |
| LOCUS      | Homo sapiens BAC clone RP11-793L24 from 2, complete sequence.       |             |     |        |                 |
| DEFINITION | ACI04634  |             |     |        |                 |
| ACCESSION  | ACI04634.5  | GI:22002223 |     |        |                 |
| VERSION    |   |             |     |        |                 |
| KEYWORDS   | HTG.  |             |     |        |                 |
| SOURCE     | human.  |             |     |        |                 |
| ORGANISM   | Homo sapiens  |             |     |        |                 |
| REFERENCE  | Fukuyama; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;    |             |     |        |                 |
| AUTHORS    | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.         |             |     |        |                 |
| TITLE      | 1 (bases 1 to 115758)   |             |     |        |                 |
| JOURNAL    | Sulston,J.E. and Waterston,R.                                       |             |     |        |                 |
| REFERENCE  | Toward a complete human genome sequence                             |             |     |        |                 |
| AUTHORS    | genome Res. 8 (11), 1097-1108 (1998)                                |             |     |        |                 |
| TITLE      | 99063792  |             |     |        |                 |
| JOURNAL    | 2 (bases 1 to 115758)   |             |     |        |                 |
| REFERENCE  | Du,H., Kozlowski,A. and Spalding,L.                                 |             |     |        |                 |
| AUTHORS    | The sequence of Homo sapiens BAC clone RP11-793L24                  |             |     |        |                 |
| TITLE      | Unpublished (2001)  |             |     |        |                 |
| JOURNAL    | 3 (bases 1 to 115758)   |             |     |        |                 |
| REFERENCE  | Waterston,R.H.  |             |     |        |                 |
| AUTHORS    | Direct Submission   |             |     |        |                 |
| TITLE      | Submitted (15-DEC-2001) Genome Sequencing Center, Washington        |             |     |        |                 |
| JOURNAL    | University School of Medicine, 4444 Forest Park Parkway, St. Louis, |             |     |        |                 |
| REFERENCE  | MO 63108, USA   |             |     |        |                 |
| AUTHORS    | 4 (bases 1 to 115758)   |             |     |        |                 |
| TITLE      | Waterston,R.H.  |             |     |        |                 |
| JOURNAL    | Direct Submission   |             |     |        |                 |
| REFERENCE  | Submitted (21-FEB-2002) Genome Sequencing Center, Washington        |             |     |        |                 |
| AUTHORS    | University School of Medicine, 4444 Forest Park Parkway, St. Louis, |             |     |        |                 |
| TITLE      | MO 63108, USA   |             |     |        |                 |
| JOURNAL    | 5 (bases 1 to 115758)   |             |     |        |                 |
| REFERENCE  | Waterston,R.H.  |             |     |        |                 |
| AUTHORS    | Direct Submission   |             |     |        |                 |
| TITLE      | Submitted (29-JUL-2002) Genome Sequencing Center, Washington        |             |     |        |                 |
| JOURNAL    | University School of Medicine, 4444 Forest Park Parkway, St. Louis, |             |     |        |                 |
| REFERENCE  | MO 63108, USA   |             |     |        |                 |
| AUTHORS    | 6 (bases 1 to 115758)   |             |     |        |                 |
| TITLE      | Waterston,R.  |             |     |        |                 |
| JOURNAL    | Direct Submission   |             |     |        |                 |
| REFERENCE  | Submitted (10-AUG-2002) Department of Genetics, Washington          |             |     |        |                 |
| AUTHORS    | University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA |             |     |        |                 |
| TITLE      | On Jul 29, 2002 this sequence version replaced gi:18855170.         |             |     |        |                 |
| JOURNAL    | -- Genome Center  |             |     |        |                 |
| COMMENT    | Center: Washington University Genome Sequencing Center              |             |     |        |                 |
|            | Center code: WUGSC  |             |     |        |                 |
|            | Web site: http://genome.wustl.edu/gsc                               |             |     |        |                 |
|            | Contact: saplens@watsn.wustl.edu                                    |             |     |        |                 |
|            | -- Summary Statistics   |             |     |        |                 |
|            | Center project name: H_NH0793L24                                    |             |     |        |                 |
|            | --  |             |     |        |                 |
|            | NOTICE: This sequence may not represent the entire insert of this   |             |     |        |                 |
|            | clone. It may be shorter because we only sequence overlapping       |             |     |        |                 |
|            | clone sections once, or longer because we provide a small overlap   |             |     |        |                 |
|            | between neighboring data submissions.                               |             |     |        |                 |
|            | This sequence was finished as follows unless otherwise noted:       |             |     |        |                 |
|            | all regions were double stranded, sequenced with an alternate       |             |     |        |                 |
|            | chemistry or covered by high quality data (i.e., phred quality >    |             |     |        |                 |
|            | 30); an attempt was made to resolve all sequencing problems, such   |             |     |        |                 |
|            | as compressions and repeats; all regions were covered by sequence   |             |     |        |                 |
|            | from more than one subclone; and the assembly was confirmed by      |             |     |        |                 |
|            | restriction digest.   |             |     |        |                 |
|            | MAPPING INFORMATION:  |             |     |        |                 |
|            | Mapping information for this clone was provided by Dr. John D.      |             |     |        |                 |
|            | McPherson, Department of Genetics, Washington University, St. Louis |             |     |        |                 |
|            | MO. For additional information about the map position of this       |             |     |        |                 |
|            | sequence, see http://genome.wustl.edu/gsc                           |             |     |        |                 |

**SOURCE INFORMATION:** The RPEC1-11 human BAC library was made from the blood of one male donor, as described by Oseogawa, K., Moon, P.-Y., Zhao, B.-Y., Frengen, E., Tateno, M., Cattanese, J.J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:11-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at <http://www.chori.org>

**NEIGHBORING SEQUENCE INFORMATION:**  
The clone sequenced to the left is Rp11-56215, 2000 bp overlap; the clone sequenced to the right is Rp11-690121, 2000 bp overlap. Actual start of this clone is at base position 131499 of Rp11-56215; actual end is at base position 65545 of Rp11-690121.

Simple sequence repeats exist between 20489 and 21499. Polymorphisms exist between AC017104, AC073476 and AC104634.

| FEATURES      |  | Location/Qualifiers   |
|---------------|--|---|
| source        |  | 1..115758   |
| repeat_region |  | /organism="Homo sapiens"<br>/db_xref="taxon:9606<br>/chromosome="2"<br>/map="2" |
| repeat_region |  | /clone="RP11-793424"<br>/clone_id="RP11-793424"                                 |
| repeat_region |  | 1..280  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | /rpt_family="(7AAA)n"   |
| repeat_region |  | 289..320  |
| repeat_region |  | 335..634  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 1043..1348  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 1356..1489  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 2044..2092  |
| repeat_region |  | /rpt_family="(7AA)n"  |
| repeat_region |  | 2094..2424  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 2427..2605  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 2609..2676  |
| repeat_region |  | /rpt_family="MLR"   |
| repeat_region |  | 2759..2790  |
| repeat_region |  | /rpt_family="AT-rich"   |
| repeat_region |  | 2791..3080  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 3083..3378  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 3441..3564  |
| repeat_region |  | /rpt_family="L1"  |
| repeat_region |  | 3602..3643  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 3645..3838  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 3899..4051  |
| repeat_region |  | /rpt_family="L1"  |
| repeat_region |  | 4300..4589  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 4686..4846  |
| repeat_region |  | /rpt_family="L1"  |
| repeat_region |  | 5660..5955  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 6048..6347  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 6383..6447  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 6448..6737  |
| repeat_region |  | /rpt_family="MLR"   |
| repeat_region |  | 6743..6778  |
| repeat_region |  | /rpt_family="Alu"   |
| repeat_region |  | 6779..6807  |
| repeat_region |  | /rpt_family="(7TA)n"  |

```
repeat_region      /rpt_family="MaLR"
                    6808..6843
repeat_region      /rpt_family="ERV1"
                    6844..7149
repeat_region      /rpt_family="Alu"
                    7150..7382
repeat_region      /rpt_family="ERV1"
                    7383..7681
repeat_region      /rpt_family="Alu"
                    7682..7913
repeat_region      /rpt_family="ERV1"
                    7914..8310
repeat_region      /rpt_family="MaLR"
                    8315..8625
repeat_region      /rpt_family="Alu"
                    8620..9199
repeat_region      /rpt_family="MaLR"
                    9205..9366
repeat_region      /rpt_family="MER103"
                    9371..9686
repeat_region      /rpt_family="Alu"
                    9703..9875
repeat_region      /rpt_family="MaLR"
                    11178..11536
repeat_region      /rpt_family="ERV1"
                    11839..11880
repeat_region      /rpt_family="ERV1"
                    11881..12178
repeat_region      /rpt_family="Alu"
                    12179..12419
repeat_region      /rpt_family="ERV1"
                    12420..12732
repeat_region      /rpt_family="Alu"
                    12723..12876
repeat_region      /rpt_family="ERV1"
                    12877..13173
repeat_region      /rpt_family="Alu"
                    13174..13641
repeat_region      /rpt_family="ERV1"
                    14200..14389
repeat_region      /rpt_family="L1"
                    14403..14429
repeat_region      /rpt_family="(T)n"
                    14466..14776
repeat_region      /rpt_family="Alu"
                    14825..14878
repeat_region      /rpt_family="(TC)n"
                    16155..16213
repeat_region      /rpt_family="MIR"
                    16214..16523
repeat_region      /rpt_family="Alu"
                    16625..16848
repeat_region      /rpt_family="MaLR"
                    16854..17157
repeat_region      /rpt_family="Alu"
                    17165..17277
repeat_region      /rpt_family="MaLR"
```

Query Match 5.98; Score 88.4; DB 9; Length 115758;

Best Local Similarity 43.48; Pred. No. 1e-11; Matches 398; Conservative 1; Mismatches 517; Indels 0; Gaps 0;

```
QY 1 GAVGGGAGCAGTGTACGATGGGCAAGCAAGATGACACCATGACACCAACGAT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21399 GATGCTGTGCTGAGAGTACTGATGATGATGATGATGATGATGATGATGAT 21340
QY 61 GATACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21339 GATGCTAATGCTGATGCTGATGCTGAAGATGACAGTATGATGATGATGATGAT 21280
QY 121 GACCAAAACCTCATATGCTACACCTTCACCAACATGTCATGCGAGCATAGAGTTG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21279 AATGATGTATGCTGATGCTGTGCTGCTGATGATGATGATGATGATGATGATGAT 21220
```

```
QY 181 TCACAGAGGAGTCATGAGACCTGTTATCTAGAACCTTCATCTGCGATTCAACAAAGT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21219 AATATATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21160
QY 241 GAAGACCATGACGACACCATGATGATGATGATGATGATGATGATGATGATGATGAT 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21159 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21100
QY 301 GCTTGTGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21099 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21040
QY 361 GACCTGGTGACCTGGTGTGACATGATAGGGCGCTGTTAATGAATTCATATATGCT 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 21039 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20980
QY 421 TGGTTGACATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20979 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20920
QY 481 ATTTTCAAGGAGTCACACCATGCTGATACCCACGATGAATCCCTGTTGTT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20919 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20860
QY 541 ATTGTCATGAGAAAGCTGCGCCAGAAACAGCAGCTGCTACATCAGAGTAGAGAA 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20859 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20800
QY 601 GATAAACAGGACATTTATGCCATTTGATGACCTGAAGATCTTAATACACACCAACAGCT 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20799 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20740
QY 661 CTCATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20739 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20680
QY 721 GTGTCCATCATTTAGAGGATTTAATGATGATGATGATGATGATGATGATGATGAT 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20679 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20620
QY 781 GTACAGATCTACGCCACGGTGACCTGACACAGTGTGATTAACCTCGGTCTAATAT 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20619 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20560
QY 841 GATCCTCATGAAAGATTACCAAGTGAGTTGGGTATGAGAGATTAATTCACGATGATGAC 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20559 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 20500
QY 901 CATGGCTTGTCTCAATG 916
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 20499 GATGGTATGGGGATG 20484
```

## RESULT 3

AC079176

LOCUS

Homo sapiens X BAC Rpl1-325Ds (Roswell Park Cancer Institute Human

DEFINITION AC079176 complete sequence.

AC079176.15 GI:21629127

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Human sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 186278)  
Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,  
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarella,J.,  
Beneton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Buhalay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Drepper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, P., Garza, N., Gill, R., Correll, J.H., Guevara, W., Guadalupe, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hayes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homati, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, T., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovach, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichteage, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., Moleod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokweto, S., Ogutu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolle, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Tansey, J., Taylor, C., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vazquez, L., Yeera, V., Villalobos, D., Winsen, R., Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Woodén, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 186278)  
Worley, K.C.

Direct Submission  
Submitted (23-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One 3 (bases 1 to 186278)  
Worley, K.C.

Direct Submission  
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 29, 2002 this sequence version replaced g1:13236484.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc.help@bcm.tmc.edu](mailto:gc.help@bcm.tmc.edu)

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:  
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2

reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: [http://www.hgsc.bcm.tmc.edu:8088/quality\\_info/genbank.annot](http://www.hgsc.bcm.tmc.edu:8088/quality_info/genbank.annot)

#### FEATURES

| source        | Location/Qualifiers  |
|---------------|--|
| repeat_region | 1..186278<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/chromosome="X"<br>/clone="RP11-325D5" |
| repeat_region | 195..244<br>/rpt_family="(CA)n"  |
| repeat_region | 1562..1592<br>/rpt_family="(GA)n"  |
| repeat_region | 1862..1939<br>/rpt_family="(C-rich)"   |
| repeat_region | 2060..2135<br>/rpt_family="(GC-rich)"  |
| repeat_region | 2293..2389<br>/rpt_family="(GC-rich)"  |
| repeat_region | complement(5038..5131)<br>/rpt_family="MIR"  |
| repeat_region | complement(5883..6055)<br>/rpt_family="MIR"  |
| repeat_region | 6132..6163<br>/rpt_family="(TG)n"  |
| repeat_region | complement(6344..6647)<br>/rpt_family="(Alu)X"   |
| repeat_region | 7051..7173<br>/rpt_family="(FLAM-C)"   |
| repeat_region | 7188..7482<br>/rpt_family="(Alu)X"   |
| repeat_region | 9120..9424<br>/rpt_family="(Alu)X"   |
| repeat_region | complement(9835..10137)<br>/rpt_family="(Alu)X"  |
| repeat_region | 10355..10578<br>/rpt_family="(Alu)X"   |
| repeat_region | complement(10583..10628)<br>/rpt_family="(Alu)X"   |
| repeat_region | complement(10629..10822)<br>/rpt_family="(Alu)X"   |
| repeat_region | complement(10852..11168)<br>/rpt_family="(Alu)X"   |
| repeat_region | 11349..11473<br>/rpt_family="(MER3)"   |
| repeat_region | complement(11512..11741)<br>/rpt_family="(MIR1B)"  |
| repeat_region | complement(11742..12124)<br>/rpt_family="(MIR1B)"  |
| repeat_region | 12505..12536<br>/rpt_family="(CG)n"  |
| repeat_region | 12536..12568<br>/rpt_family="(CG)n"  |
| repeat_region | complement(12672..12961)<br>/rpt_family="(Alu)X"   |
| repeat_region | complement(13662..13970)<br>/rpt_family="(Alu)X"   |
| repeat_region | 14644..14944<br>/rpt_family="(Alu)X"   |
| repeat_region | 15076..15137<br>/rpt_family="(TA)n"  |
| repeat_region | 15232..15407<br>/rpt_family="(TA)n"  |

```

repeat__region      15427. .15485
/rpt_family="AT_rich"
repeat__region      15513. .15598
/rpt_family="(TA)n"
repeat__region      15599. .15892
/rpt_family="AluSp"
STS                 15688. .15803
/standard_name="D11S2560"
repeat__region      15893. .15979
/rpt_family="(TA)n"
repeat__region      16010. .16170
/rpt_family="(TA)n"
repeat__region      complement(16224. .16405)
/rpt_family="MER33"
repeat__region      16406. .16832
/rpt_family="MSB"
repeat__region      complement(16833. .16881)
/rpt_family="MER33"
repeat__region      17271. .17567
/rpt_family="AluY"
repeat__region      complement(17628. .17781)
/rpt_family="L1MC3"
repeat__region      complement(17816. .17866)
/rpt_family="MSB"
repeat__region      complement(17961. .18065)
/rpt_family="MSB"
repeat__region      complement(18068. .18152)
/rpt_family="L1MC3"
repeat__region      18187. .19078
/rpt_family="AluYo"
repeat__region      19082. .19165
/rpt_family="AluYo/FRAM"
STS                 20379. .20480
/standard_name="STSG3076"
repeat__region      20690. .20723
/rpt_family="(TTA)n"
repeat__region      complement(20724. .21004)

```

[illegible][illegible]

|            |  |
|------------|--|
| RESULT 4   |  |
| LOCUS      | AC090533   |
| DEFINITION | Mus musculus clone RP2-33H13 strain C57BL6/J, WORKING DRAFT  |
| ACCESSION  | AC090533   |
| VERSION    | AC090533.4 GI:1976439  |
| KEYWORDS   | HTG; HTGS_PHASE1; HTGS_DRAFT.  |
| SOURCE     | Mus musculus.  |
| ORGANISM   | Fungiarya: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.   |
| REFERENCE  | 1 (bases 1 to 272545)  |
| AUTHORS    | I.L., Montgomery,K.T., Gills,G., Chu,D., Decker,J., Fusina,M.,<br>Goltz,J., Halder,A., Hall,L., Han,T., Ioshikhes,I.P., Lee,E.,<br>Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.<br>High Throughput Mouse Sequencing<br>Unpublished<br>2 (bases 1 to 272545)   |
| TITLE      | I.L., Montgomery,K.T., Gills,G., Chu,D., Decker,J., Fusina,M.,<br>Goltz,J., Halder,A., Hall,L., Han,T., Ioshikhes,I.P., Lee,E.,<br>Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.<br>Direct Submission<br>Submitted (02-Mar-2001) Department of Molecular Genetics, Albert<br>Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,<br>Bronx, NY 10461, USA<br>On Dec 21, 2001 this sequence version replaced gi:13310871. |
| JOURNAL    | -----Genome Center   |
| COMMENT    | Center: Harvard Partners Genome Center<br>Web site: http://www.hpcg.org/sequence/mouse.html<br>Contact: hpgc@med.nyu.edu harvard.edu<br>-----Summary Statistics<br>Center project name: ADU<br>Sequencing vector: pUC18; L08752<br>Chemistry: Dye-terminator Big Dye; 100%<br>*Consensus quality: 257441 at least Q20<br>*Consensus quality: 253718 at least Q30<br>*Consensus quality: 247095 at least Q40                                |











\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 30267: contig of 30267 bp in length  
\* 30268 30287: gap of unknown length  
\* 30288 58400: contig of 2813 bp in length  
\* 58401 58420: gap of unknown length  
\* 58421 88864: contig of 3044 bp in length  
\* 88865 88884: gap of unknown length  
\* 88885 109889: contig of 21005 bp in length  
\* 109890 109909: gap of unknown length  
\* 109910 125133: contig of 15224 bp in length  
\* 125134 125153: gap of unknown length  
\* 125154 139549: contig of 14386 bp in length  
\* 139550 139569: gap of unknown length  
\* 139570 153410: contig of 13841 bp in length  
\* 153411 153430: gap of unknown length  
\* 153431 162567: contig of 9137 bp in length  
\* 162568 162587: gap of unknown length  
\* 162588 171044: contig of 8457 bp in length  
\* 171045 171065 179304: contig of 8240 bp in length  
\* 179305 179324: gap of unknown length  
\* 179325 187059: contig of 7735 bp in length  
\* 187060 187079: gap of unknown length  
\* 187080 197598: contig of 10539 bp in length  
\* 197599 197618: gap of unknown length  
\* 197619 205986: contig of 8348 bp in length  
\* 205987 205989: gap of unknown length  
\* 205989 212449: contig of 6463 bp in length  
\* 212450 212469: gap of unknown length  
\* 212470 217752: contig of 5283 bp in length  
\* 217753 217772: gap of unknown length  
\* 217773 223943: contig of 6171 bp in length  
\* 223944 223963: gap of unknown length  
\* 223964 230362: contig of 6399 bp in length  
\* 230363 230382: gap of unknown length  
\* 230383 236251: contig of 5869 bp in length  
\* 236252 236271: gap of unknown length  
\* 236272 240572: contig of 4301 bp in length  
\* 240573 240592: gap of unknown length  
\* 240593 242789: contig of 2197 bp in length  
\* 242790 242809: gap of unknown length  
\* 244695 244695: contig of 1886 bp in length  
\* 244696 244715: gap of unknown length  
\* 244716 246969: contig of 2254 bp in length  
\* 246970 247870: gap of unknown length  
\* 247871 247890: contig of 881 bp in length  
\* 247891 248193: gap of unknown length  
\* 248194 248213: contig of 303 bp in length  
\* 248214 250122: gap of unknown length  
\* 250123 250142: contig of 1909 bp in length  
\* 250143 251367: contig of 1225 bp in length  
\* 251368 251387: gap of unknown length  
\* 251388 252749: contig of 1362 bp in length  
\* 252750 252769: gap of unknown length  
\* 252770 253640: contig of 871 bp in length  
\* 253641 253660: gap of unknown length  
\* 253661 255032: contig of 1372 bp in length  
\* 255033 255052: gap of unknown length  
\* 255053 257333: contig of 2281 bp in length  
\* 257334 257353: gap of unknown length  
\* 257354 259409: contig of 2056 bp in length  
\* 259410 259429: gap of unknown length  
\* 259430 260459: contig of 1030 bp in length  
\* 260460 260479: gap of unknown length  
\* 260480 260829: contig of 350 bp in length  
\* 260830 260849: gap of unknown length  
\* 261651 261662: contig of 812 bp in length  
\* 261663 261682: gap of unknown length  
\* 261683 262532: contig of 851 bp in length

FEATURES  
source

\* 262533 262552: gap of unknown length  
\* 262553 263351: contig of 799 bp in length  
\* 263352 263371: gap of unknown length  
\* 263372 264217: contig of 846 bp in length  
\* 264218 264237: gap of unknown length  
\* 264238 265205: contig of 968 bp in length  
\* 265206 265225: gap of unknown length  
\* 265226 265270: contig of 1045 bp in length  
\* 265271 266290: gap of unknown length  
\* 266291 266749: contig of 439 bp in length  
\* 266750 266769: gap of unknown length  
\* 266770 267118: contig of 949 bp in length  
\* 267119 267738: gap of unknown length  
\* 267739 268576: contig of 838 bp in length  
\* 268577 268596: gap of unknown length  
\* 268597 270010: contig of 1414 bp in length  
\* 270011 270030: gap of unknown length  
\* 270031 271654: contig of 1624 bp in length  
\* 271655 271674: gap of unknown length  
\* 271675 272545: contig of 871 bp in length.

Location/Qualifiers  
1. -272545  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RP23-33H13"  
/sex="male"  
1. 30267  
/note="assembly\_name:Contig108"  
30288. -58400  
/note="assembly\_name:Contig107  
clone\_end:SP6  
vector\_side:left"  
58421. -88864  
/note="assembly\_name:Contig106"  
88885. -109889  
/note="assembly\_name:Contig105"  
109910. -125133  
/note="assembly\_name:Contig104  
clone\_end:T7  
vector\_side:right"  
125154. -139549  
/note="assembly\_name:Contig103"  
139570. -153410  
/note="assembly\_name:Contig102"  
153431. -162567  
/note="assembly\_name:Contig101"  
162588. -171044  
/note="assembly\_name:Contig100"  
171065. -179304  
/note="assembly\_name:Contig99"  
179325. -187059  
/note="assembly\_name:Contig98"  
187080. -197598  
/note="assembly\_name:Contig97"  
197619. -205966  
/note="assembly\_name:Contig96"  
205987. -212449  
/note="assembly\_name:Contig95"  
212470. -217752  
/note="assembly\_name:Contig94"  
217773. -223943  
/note="assembly\_name:Contig93"  
223964. -230362  
/note="assembly\_name:Contig92"  
230383. -236251  
/note="assembly\_name:Contig91"  
236272. -240572  
/note="assembly\_name:Contig90"  
240593. -242789  
/note="assembly\_name:Contig89"  
242810. -244695  
/note="assembly\_name:Contig88"



|            |   |   |                            |
|------------|---|---|----------------------------|
| Db         | 157652  | ATGATGTTGATCTTATTTATGCGTATGATGCATATGATGATGATGATGATGATTAATGTTTA    | 157593                     |
| Oy         | 479   | CTAT-----TTTACAAGGAATCACCACCATGATACCCACGACAGTAATG                 | 529                        |
| Db         | 157592  | ATAATGATGCGTATTTATGATGACTAAGATTAACGATTAATGATGATGATGATGACGATG      | 157533                     |
| Oy         | 530   | CCGTGTCGTATTTGGTCATGGAAGAAAGCGCCGCAAGAAACGACGCGCTCATTACACG        | 589                        |
| Db         | 157332  | ATGATGCGTATGACGATGATTAATGATGATGATGATGATGATGATGATGATGATGATG        | 157473                     |
| Oy         | 590   | ACGTAGAGGAAGTAAAACTGAGCATTTATGCCATTGTGACGTAAATATACACAC            | 649                        |
| Db         | 157472  | AGGTTGACTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG       | 157413                     |
| Oy         | 650   | MACCAAAGCCTCTTCATCATCATGTCACGCAAGCAACCATGATTTCAAAAGTGGTTATG       | 709                        |
| Db         | 157412  | ATGCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 157353                     |
| Oy         | 710   | GTGACCTTGAAGTGTCTCTACACATTTAGAGGATTTTAATGTAACTGATGACACACAAGATC    | 769                        |
| Db         | 157352  | GTGCGTATATGTTGATTAATGATGAAAGTGATTAATGATTAATGAAGATGATGATGAAGATG    | 157293                     |
| Oy         | 770   | ATCCTCATGCGATGACGATCTACGCCAACGGAGACCTACACGTGATGATGATGATGATGATG    | 829                        |
| Db         | 157292  | ATGCTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG    | 157233                     |
| Oy         | 830   | GTGCTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG   | 889                        |
| Db         | 157232  | ATGTTTATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 157173                     |
| Oy         | 890   | ACGATGATGACCATGCGCTTGTCATGAA 919                                  |                            |
| Db         | 157172  | ACTATG   | 157143                     |
| RESULT 9   |   |   |                            |
| LOCUS      | AC018695/c  | 211001 bp   | DNA linear PRI 09-JUN-2002 |
| DEFINITION | Homo sapiens BAC clone Rpl1-568023 from 16, complete sequence.      |   |                            |
| ACCESSION  | AC018695  |   |                            |
| VERSION    | AC018695.6  | GI:13443271   |                            |
| KEYWORDS   | HTG.  |   |                            |
| SOURCE     | Homo sapiens.   |   |                            |
| ORGANISM   | Homo sapiens.   |   |                            |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  |   |                            |
| AUTHORS    | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.           |   |                            |
| TITLE      | 1 (bases 1 to 211001)   |   |                            |
| JOURNAL    | Sulston,J.E. and Watson,R.  |   |                            |
| MEDLINE    | Toward a complete human genome sequence                             |   |                            |
| PUBMED     | Genome Res. 8 (11), 1097-1108 (1998)                                |   |                            |
| REFERENCE  | 99063792  |   |                            |
| AUTHORS    | 2 (bases 1 to 211001)   |   |                            |
| TITLE      | Hakenson,W., Scott,K., Courtney,L., Drone,K., Gregory,S. and        |   |                            |
| JOURNAL    | Lestly,K.   |   |                            |
| REFERENCE  | The sequence of Homo sapiens BAC clone Rpl1-568023                  |   |                            |
| AUTHORS    | Unpublished (2001)  |   |                            |
| TITLE      | 3 (bases 1 to 211001)   |   |                            |
| JOURNAL    | Waterson,R.H.   |   |                            |
| REFERENCE  | Direct Submission   |   |                            |
| AUTHORS    | Submitted (16-Dec-1999) Genome Sequencing Center, Washington        |   |                            |
| TITLE      | University School of Medicine, 4444 Forest Park Parkway, St. Louis, |   |                            |
| JOURNAL    | MO 63108, USA   |   |                            |
| REFERENCE  | 4 (bases 1 to 211001)   |   |                            |
| AUTHORS    | Waterson,R.H.   |   |                            |
| TITLE      | Direct Submission   |   |                            |
| JOURNAL    | Submitted (24-MAR-2001) Genome Sequencing Center, Washington        |   |                            |
| REFERENCE  | University School of Medicine, 4444 Forest Park Parkway, St. Louis, |   |                            |
| AUTHORS    | MO 63108, USA   |   |                            |
| TITLE      | 5 (bases 1 to 211001)   |   |                            |
| JOURNAL    | Waterston,R.  |   |                            |
| REFERENCE  | Direct Submission   |   |                            |

```

JOURNAL
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
6 (bases 1 to 211001)
AUTHORS
Waterston, R.
TITLE
Direct Submission
JOURNAL
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
On Mar 24, 2001 this sequence version replaced gi:9838280.
-----
Center: Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
Summary Statistics
Center project name: H_NH0568J23
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
Mo. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPci-11 human BAC library was made from the blood of one male
donor, as described by Osoegewa, K., Moon, P. T., Zhao, B., Frengen, E.,
Tateno, H., Catalanese, J. J. and de Jong, P. J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-542M13. Actual start of
this clone is at base position 1 of RP11-568J23; actual end is at
base position 211001 of RP11-568J23.

H_NH0568J23 contains an imperfect GA run from 143896 to 144293, in
which the exact length is unknown. There are approximately 800
bases missing according to the restriction digests and per data.

H_NH0568J23 contains a single stranded region from 144079 to 144172
which contains low quality data.
Location/Qualifiers
1. 211001
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16"
/clone="RP11-568J23"
/clone_lib="RPci-11"
354..361
/ftc_family="Alu"
407..467
/ftc_family="MIR"
760..835
/ftc_family="MIR"
891..929
/ftc_family="MER1_type"

```



|              |  |  |
|--------------|--|--|
|              | SOURCE   | Dicystostellium discoideum.  |
|              | ORGANISM   | Dicystostellium discoidem.   |
|              | REFERENCE  | Eukaryota; Mycetozoa; Dicystostellida; Dicystostellium.<br>1 (bases 1 to 61052)  |
|              | AUTHORS  | Glockner, G., Eichinger, L., Szafanski, K., Pachbat, J., Dear, P.,<br>Lehmann, R., Baumgart, C., Parra, G., April, J.F., Guigo, R., Kumpf, K.,<br>Tunggal, B., Cox, E., Quail, M.A., Platzer, W., Rosenthal, A. and<br>Noegel, A.A.  |
| TITLE        | Title  | Unpublished Sequence and Analysis of Chromosome 2 of Dicystostellium   |
| JOURNAL      | The Dicystostellium Genome Sequencing Consortium                     | 2 (bases 1 to 61052)<br>Baumgart, C.   |
| REMARK       | Direct Submission  | Submitted (06-APR-2002) Genome Analysts, Institute of Molecular<br>Biotechnology, Beutenbergstr. 11, Jena 07745, Germany   |
| REFERENCE    | CDS predictions from GenBank may contain errors. Further Information | is available from IMB Jena, Department of Genome Analysis<br>(http://genome.imb-jena.de/dicystostellium/<br>and the University Cologne, Institute for Project.shtml<br><a href="http://www.uni-koeeln.de/dicystostellium/project.shtml">http://www.uni-koeeln.de/dicystostellium/project.shtml</a> ) |
| COMMENT      | Funding Agency :   | Deutsche Forschungsgemeinschaft (DFG).   |
|              | * NOTE:  | This is a 'working draft' sequence.  |
|              | *  | This sequence will be replaced   |
|              | *  | by the finished sequence as soon as it is available and  |
|              | *  | the accession number will be preserved.  |
| FEATURES     | Location/Qualifiers  | 1..61052 "Dicystostellium discoidem"<br>organism=<br>"/strain='Xk4'<br>'/Db.Xref='Taxon:44689'<br>'/Chromosome='2'<br>'/map='2779865-2840915'<br>BASE COUNT      25445    a    6805    c    6328    g    22474    t<br>ORIGIN  |
| Query Match  | 3.9%; Score 58.4;  | DB 2; Length 61052;<br>Best Local Similarity 45.7% Pred. No. 0.00067;  |
| Matches 200; | Conservative 1;  | Mismatches 237; Indels 0; Gaps 0;  |
| OY           | 1  | GAGGGGACGAGGTGAATCATTGCAGCAAAAGAATGCACCATCGAACCCACCAT 60   |
| DB           | 7703   | GATGATGAGATGAGATGAGATGAGTGATGAGTTCCTTTATCTTAGAATCTTA 111   |
| OY           | 61   | GATCACCATGACGCCATGATGATGATGATAATGAACAATGACTTATGCCCATGTGAATG 120  |
| DB           | 7763   | GAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 7822  |
| OY           | 121  | GAACCAAACCTTATATGCTTACGACACCTTACCAACCAATCTTCATTCGACAGATGAGTTG 180  |
| DB           | 7823   | GAT 7882  |
| OY           | 181  | TCACAGAGGGTCATGAGACCTGTTTTATCTTAGAATCTTATCTTTGTCGATTCACACAGT 240   |
| DB           | 7883   | GAT 7942  |
| OY           | 241  | GAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300   |
| DB           | 7943   | GAT 8002  |
| OY           | 301  | GGTTGATTTCTTWTGGCGACATGTRACATGCTGCCAAGAAAACATGCTGACCCCTGGT 360   |
| DB           | 8003   | GAT 8062  |
| OY           | 361  | GACTCGGTGATCTGGTTGACGATGATGAGGGCCGTTAATGAATTCATCATTTATGCT 420  |
| DB           | 8063   | GAT 8122  |
| OY           | 421  | TGCTTGACATTTGATGCT 438   |
| DB           | 8123   | TATTTGATTTGATGATGAT 8140   |

|            |   |          |     |        |                 |
|------------|---|----------|-----|--------|-----------------|
|            | ACI123513   | 61052 bp | DNA | linear | HTG-29-MAY-2002 |
| RESULT 11  |   |          |     |        |                 |
| LOCUS      | ACI123513   |          |     |        |                 |
| DEFINITION | Dictyostellium discoideum chromosome 2 map 2779865-2840915 strain AX4 *** SEQUENCING IN PROGRESS ***, in ordered pieces.  |          |     |        |                 |
| ACCESSION  | ACI123513   |          |     |        |                 |
| VERSION    | ACI123513.1 GI:21240650   |          |     |        |                 |
| KEYWORDS   | HTG; HTGS_PHASE2  |          |     |        |                 |
| SOURCE     | Dictyostellium discoideum.  |          |     |        |                 |
| ORGANISM   | Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.  |          |     |        |                 |
| REFERENCE  | 1 (bases 1 to 61052)<br>Gloekner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P., Lehmann,R., Baumgart,C., Parra,G., April,I.F., Guigo,R., Kumpf,K., Junggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and Noegel,A.A.<br><i>Sequence and Analysis of Chromosome 2 of Dictyostellium</i>   |          |     |        |                 |
| AUTHORS    | Unpublished<br>The Dictyostellium Genome Sequencing Consortium<br>2 (bases 1 to 61052)  |          |     |        |                 |
| TITLE      | Baumgart,C.   |          |     |        |                 |
| REMARK     | Direct Submission   |          |     |        |                 |
| REFERENCE  | Submitted (29-May-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany   |          |     |        |                 |
| AUTHORS    | CDS predictions from Genemid may contain errors. Further Information is available from IMB-Jena, Department of Genome Analysis ( <a href="http://genome.imb-jena.de/dictyostellium/">http://genome.imb-jena.de/dictyostellium/</a> ) and the University Cologne, Institute for Biochemistry I ( <a href="http://www.uni-koeln.de/dictyostellium/project.shtml">http://www.uni-koeln.de/dictyostellium/project.shtml</a> )<br>Funding Agency : Deutsche Forschungsgemeinschaft (DFG).  |          |     |        |                 |
| JOURNAL    | * NOTE: This is a working draft sequence.<br>* This sequence will be replaced<br>* by the finished sequence as soon as it is available and<br>* the accession number will be preserved.<br>Location/Qualifiers<br>1..61052  |          |     |        |                 |
| COMMENT    | location= "dictyostellium discoideum"<br>organism="Dictyostellium discoideum"<br>strain="AX4"<br>db_xref="taxon:44689"<br>chromosome="2"<br>map="2779865-2840915"<br>join(448..628,705..803,875..1134)<br>note="ORF_ID:dd_02708"  |          |     |        |                 |
| FEATURES   | pseudo  |          |     |        |                 |
| source     | codon_start=1<br>complement(join(1193..1583,1653..6117,6198..6330))<br>note="ORF_ID:dd_02705"   |          |     |        |                 |
| CDS        | /codon_start=1<br>/cdoen_start=1<br>/product="ABC1 TRANSPORTER"<br>protein_id="AAK44362.1"<br>db_xref="GI:21240651"<br>translation="MGERSQLTKLTLLKNLKLKSRCGCCICEVPIIIIVLFAAILVLVQFDYDLIKTQFSNRIENNNIYYGGAGSLANEOKGVADMMKPOLSNLEFKSISEVDVFEKEINDRSEMENVFOINSTOVNGEIPESNDLSAVNTMTNTTSFYXSTRDSSNVLDNSKRKWDNGIDSDSYFLAKMAYIQIMDADEALFGYGDKRYKLVNGOXPDPPEYLMOVKMINGRESVEPKASGVSAALQJMLMDAIEGYGLDYKVIENGOPYLYNQVFISWITSLVALPILDIILIIYIKSGSVIHSTEWMIIVITLLIYLPOLLALAFFSWEPDKSRFPAGLIPLFIILLINCIGIFGEIEINHILKLCICFCSPDIGAGCFYMSFRDIPTDISTINTLNVOYSKEOIIGTFEVPIIFYPLMYLDKIYKTGGRKBMYELFKRWTVWGAKRKNNKKNEKYENIDEESTIONNNDNJEMPJEVRKTTISRNRKFKTGDELRTIAVNDNLDMEDDOHIGLGPNKGSCSTGISLGTGSTGAFLNIGNTDTOMNEIRKCTGCYCLONDIYMNOYLIVHEHYIAISKITTNRKOIKRSKLMAANEIDIGKITTPAGCSISGCKRRKICGATGFTRSTFTPDVSGSDSSRRVPMLSHSKGKTILTTHRIDEDADISDRISTISHKRLTDGSSLFLNKRGVGYLTLCKSLANTIDQVNADVOTQFTRDOIPLYVTASNNGSISIRLCPYAPLFVSFTPODBENLKSNHIDYSIVSTILLEEFALKIGTDIPDALISITSLRPLEVFNKSTEGSLKSOOLKALLIKRVKTSKDFKAFPLLILSLPLLAIGSIIYFREVDEVETFINNSTEPLTESLLQOYKRNQIVPIQLAGTTSSEDFNRKELSNPYNOPYOYANLINENOYLANNYOSGSLINFPIPLSSLSINNTRKVYSNSLENFNITHWPVHNLDNALRHNDIGTECHSLPFHDILLSFORKASOGMNIOIYAFIWMAGSFMAGSFAGISOERTFNSIKRLAIYISOCKRYIWSVSNLMIDYFAFIFILITTCIIILAIVDDKFENDQGFEFLSVIFSLSIIP |          |     |        |                 |





QY 181 TCACAGAGGTCATGAGCTTTATCTAGACTTCTGTCGATTCAACACAGT 240  
DB 7883 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7942  
QY 241 GAAGACACGAGACACACACACACACTTCATGACATGCTTGACATGACAC 300  
DB 7943 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8002  
QY 301 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8062  
DB 8003 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8062  
QY 361 GACCTGCTGACCTGCTGACATGATGAGGCGGTGTAAGTTCATGATGCT 420  
DB 8063 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8122  
QY 421 TGGTGGACATTGATGCT 438  
DB 8123 TATTAGCATGATGAGGAT 8140

RESULT 12  
AC131509 163690 bp DNA linear HTG 23-AUG-2002  
LOCUS Strongylocentrotus purpuratus clone Sp58L24, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 7 ordered pieces.  
ACCESSION AC131509  
VERSION AC131509.1 GI:22450561  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_PULLTOP.  
SOURCE Strongylocentrotus purpuratus.  
ORGANISM Strongylocentrotus purpuratus.  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinacea; Echinoda;  
Strongylocentrotidae; Strongylocentrotus.  
REFERENCE 1 (bases 1 to 163690)  
AUTHORS Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caletani, C.,  
Yuh, C.H., Minokawa, T., Amore, G., Hinnman, V., Atenese-Mena, C.,  
Schliff, M.J., Clarke, P.J.C., Rust, A.G., Pan, Z., Arnone, M.I.,  
Rowen, L., Cameron, R., Andrew, M., McClay, D.R., Hood, L. and Bolouri, H.  
A provisional regulatory gene network for specification of  
endomesoderm in the sea urchin embryo  
Dev. Biol. 246 (1), 162-190 (2002)  
22024154  
2 (bases 1 to 163690)  
AUTHORS Rowen, L., Cameron, R.A. and Davidson, E.H.  
REFERENCE Direct Submission  
TITLE Submitted (23-AUG-2002) Multimegabase Sequencing Center, Institute  
JOURNAL for Systems Biology, 1441 N. 34th Street, Seattle, WA 98103, USA  
COMMENT ----- Genome Center Sequencing Center  
Center: Multimegabase Sequencing Center  
Center code: ISBMS  
Web site: http://www.systemsbio.org  
Contact: leetowen@systemsbio.org  
Drafting center: ISBMS  
----- Summary Statistics  
Sequencing vector: pUC18; 108752  
Chemistry: Dye-terminator Big Dye; 90% of reads  
Chemistry: Dye-terminator Big Dye; 10% of reads  
Assembly program: Phrap; version 0.990399  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 41287: contig of 41287 bp in length  
\* 41288 41387: gap of unknown length

41388 106251: contig of 64864 bp in length  
\* 106252 106351: gap of unknown length  
\* 106352 129007: contig of 22656 bp in length  
\* 129008 129107: gap of unknown length  
\* 129108 138703: contig of 9596 bp in length  
\* 138704 138803: gap of unknown length  
\* 138804 150338: contig of 11736 bp in length  
\* 150340 160458: gap of unknown length  
\* 160459 160559: gap of unknown length  
\* 160559 163690: contig of 3132 bp in length.  
Location/Qualifiers  
1. 163690  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="Sp58L24"  
/clone="1b"="Caltech Strongylocentrotus purpuratus sperm  
genomic BAC library A"  
/note="Caltech Strongylocentrotus purpuratus sperm genomic  
BAC library A"  
BASE COUNT 51095 a 28957 c 32280 g 50747 t 611 others  
ORIGIN

Query Match 3.9% Score 58.4; DB 2; Length 163690;  
Best Local Similarity 44.3%; Pred. No. 0.00077;  
Matches 239; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

QY 444 ACCAACAACCGAGCTGCTCATGACACTCATGATATTTTACAGGAGTACACCGA 503  
DB 72297 AACATGCCACACAGAAATTCCTGATTAAGTAACTATATGATGATGATG 72356  
QY 504 TCGTATATGCCAGCCAGCATGATGATGATGATGATGATGATGATGATGATG 72356  
DB 72357 TCGTATATGCCAGCCAGCATGATGATGATGATGATGATGATGATGATGATG 72416  
QY 564 AGAACAACGACGCTGCTCATGACACTCATGATATTTTACAGGAGTACACCGA 623  
DB 72417 TCGTATATGCCAGCCAGCATGATGATGATGATGATGATGATGATGATGATG 72476  
QY 624 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 683  
DB 72477 TCGTATATGCCAGCCAGCATGATGATGATGATGATGATGATGATGATGATG 72536  
QY 684 CACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 743  
DB 72537 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72596  
QY 744 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 803  
DB 72597 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72656  
QY 804 CCTGACACGATGATGATGATGATGATGATGATGATGATGATGATGATG 863  
DB 72657 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72716  
QY 864 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 923  
DB 72717 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72776  
QY 924 CAGATATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 983  
DB 72777 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72836

RESULT 13  
AC099341 51939 bp DNA linear PRI 22-FEB-2002  
LOCUS Homo sapiens BAC clone RP11-317C13 from 7, complete sequence.  
DEFINITION AC099341  
ACCESSION AC099341  
VERSION AC099341.4 GI:18464301  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
1 (bases 1 to 51939)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
2 (bases 1 to 51939)  
Vanbrunt, A., Kozlowicz, A. and Boyer, F.  
The sequence of Homo sapiens BAC clone RP11-317C13  
Unpublished (2001)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
3 (bases 1 to 51939)  
Waterston, R.H.  
Direct Submission  
Submitted (09-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
4 (bases 1 to 51939)  
Waterston, R.H.  
Direct Submission  
Submitted (01-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
5 (bases 1 to 51939)  
Waterston, R.  
Direct Submission  
Submitted (22-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Feb 1, 2002 this sequence version replaced g1.17977469.

-----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
-----  
Summary Statistics  
-----  
Center project name: H\_NH0317C13  
-----

COMMENT

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
All regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:legreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RP11-317C13 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catenease, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBac3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-620M21, 2000 bp overlap; the clone sequenced to the right is RP5-99414, 2000 bp overlap. Actual start of this clone is at base position 148410 of RP13-620M21.

There is a single plasmid subclone from 18209 to 18268. There is an unresolved simple sequence repeat from 3084 to 3812. Polymorphisms have been identified between AC104594, AC007129 and AC099341. Data from AC099649 was used to finish this clone, AC099341.

#### FEATURES

source  
1..51939  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7"  
/clone="RP11-317C13"  
/clone\_lib="RP11-11"  
Location/Qualifiers  
1..51939  
/repeat\_region  
1..362  
/rpt\_family="I1"  
489..542  
/rpt\_family="(GGCA)n"  
537..715  
/rpt\_family="I2"  
724..1094  
/rpt\_family="MALR"  
2169..2478  
/rpt\_family="Alu"  
2172..2194  
/rpt\_family="(TTA)n"  
2857..3072  
/rpt\_family="MIR"  
3095..3543  
/rpt\_family="(ATC)n"  
3562..3807  
/rpt\_family="(ATG)n"  
4215..4280  
/rpt\_family="AT-rich"  
4849..4894  
/rpt\_family="Alu"  
4895..5090  
/rpt\_family="Alu"  
5234..5369  
/rpt\_family="MIR1\_type"  
5364..5793  
/rpt\_family="MIR1\_type"  
5804..5900  
/rpt\_family="MIR1\_type"  
5906..6029  
/rpt\_family="(TA)n"  
6045..6159  
/rpt\_family="(CATATA)n"  
6161..6242  
/rpt\_family="(TA)n"  
6256..6437  
/rpt\_family="(TA)n"  
6442..6498  
/rpt\_family="(CAT)n"  
6493..6588  
/rpt\_family="MIR1\_type"  
6754..7105  
/rpt\_family="MALR"  
7626..7746  
/rpt\_family="MIR"  
8074..8175  
/rpt\_family="MIR"  
10558..10645  
/rpt\_family="I2"  
13385..13771  
/rpt\_family="ENVL"  
15142..15777  
/rpt\_family="MIR2\_type"  
15939..15966





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OK nucleic - nucleic search, using \$w model

Run on: November 29, 2002, 06:34:04 : Search time 268 Seconds  
(without alignments)  
12528.841 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gaggggagcagtgtaacga.....gctgttatagcagcgca 1491

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1          | 1490.6 | 100.0       | 1491   | 21 | AAA47150    |
| 2          | 1490.6 | 100.0       | 1491   | 21 | AAA47151    |
| 3          | 56.8   | 3.8         | 1083   | 23 | AA567245    |
| 4          | 52.6   | 3.5         | 583    | 22 | ABA50512    |
| 5          | 52.6   | 3.5         | 583    | 22 | ABA54653    |
| 6          | 52.6   | 3.5         | 583    | 22 | ABA35452    |
| 7          | 52.6   | 3.5         | 583    | 22 | AAK15834    |
| 8          | 52.6   | 3.5         | 583    | 22 | AAK42604    |
| 9          | 52.6   | 3.5         | 583    | 22 | AAI23356    |

|    |      |     |       |    |          |                    |
|----|------|-----|-------|----|----------|--------------------|
| 10 | 52.6 | 3.5 | 583   | 22 | AAI48680 | Probe #17366 used  |
| 11 | 52.6 | 3.5 | 583   | 24 | AAI08992 | Probe #8983 used t |
| 12 | 52.6 | 3.5 | 583   | 24 | AB516657 | Human genome-deriv |
| 13 | 52.6 | 3.5 | 1959  | 22 | ABA45382 | Human breast cell  |
| 14 | 52.6 | 3.5 | 1959  | 22 | ABA55972 | Human foetal liver |
| 15 | 52.6 | 3.5 | 1959  | 22 | ABA25346 | Probe #4012 for ge |
| 16 | 52.6 | 3.5 | 1959  | 22 | AAK04090 | Human brain expres |
| 17 | 52.6 | 3.5 | 1959  | 22 | AAK29571 | Human bone marrow  |
| 18 | 52.6 | 3.5 | 1959  | 22 | AAI14146 | Probe #4079 for ge |
| 19 | 52.6 | 3.5 | 1959  | 22 | AAI35531 | Probe #3985 used t |
| 20 | 52.6 | 3.5 | 1959  | 24 | AB504127 | Human genome-deriv |
| 21 | 49.2 | 3.3 | 327   | 22 | ABA71406 | Human foetal liver |
| 22 | 49.2 | 3.3 | 327   | 22 | AAK19726 | Human brain expres |
| 23 | 49.2 | 3.3 | 327   | 22 | AAK45746 | Human bone marrow  |
| 24 | 49.2 | 3.3 | 327   | 22 | AAI51672 | Probe #20358 used  |
| 25 | 49.2 | 3.3 | 327   | 24 | AB520024 | Human genome-deriv |
| 26 | 49.2 | 3.3 | 327   | 24 | AA578187 | DNA encoding novel |
| 27 | 46.8 | 3.1 | 8244  | 23 | ABA58874 | Human foetal liver |
| 28 | 46.8 | 3.1 | 456   | 22 | ABA66529 | DNA encoding novel |
| 29 | 46.2 | 3.1 | 456   | 22 | AAK07034 | Human bone marrow  |
| 30 | 46.2 | 3.1 | 456   | 22 | AAK32775 | Human bone marrow  |
| 31 | 46.2 | 3.1 | 456   | 22 | AAI38589 | Probe #7275 used t |
| 32 | 46.2 | 3.1 | 456   | 22 | AB507574 | Human genome-deriv |
| 33 | 46.2 | 3.1 | 456   | 22 | AB508114 | Protophila melanog |
| 34 | 45   | 3.0 | 15935 | 23 | AAI26289 | P. talcparum egg   |
| 35 | 44.8 | 3.0 | 574   | 22 | ABA51944 | Human foetal liver |
| 36 | 44.2 | 3.0 | 574   | 22 | AAK0231  | Human brain expres |
| 37 | 44.2 | 3.0 | 574   | 22 | AAK0231  | Human brain expres |
| 38 | 44.2 | 3.0 | 574   | 22 | AAK0231  | Human brain expres |
| 39 | 44.2 | 3.0 | 574   | 22 | AAK0231  | Human brain expres |
| 40 | 44.2 | 3.0 | 574   | 22 | AAI10301 | Probe #234 for gen |
| 41 | 44.2 | 3.0 | 574   | 22 | AAI13552 | Probe #238 used to |
| 42 | 44.2 | 3.0 | 574   | 22 | AAI00237 | Human genome-deriv |
| 43 | 44.2 | 3.0 | 574   | 24 | AB500249 | Human breast cell  |
| 44 | 44.2 | 3.0 | 669   | 22 | ABA64604 | Human foetal liver |

#### ALIGNMENTS

|          |  |                         |
|----------|--|-------------------------|
| RESULT 1 | AAA47150   | standard: DNA: 1491 BP. |
| ID       | AAA47150   |                         |
| AC       | AAA47150   |                         |
| DT       | 03-OCT-2000  | (first entry)           |
| DE       | DNA encoding a serine protease inhibitor protein.              |                         |
| KW       | Serine protease inhibitor; green-lipped mussel; anti-thrombin; |                         |
| KW       | divalent metal cation binding activity; dietary supplement;    |                         |
| KW       | anticoagulant; ss.   |                         |
| OS       | Perna canaliculus.   |                         |
| XX       | WO200039165-A1.  |                         |
| PN       | 06-JUL-2000.   |                         |
| PD       | 23-DEC-1999;   | 99WO-NZ00227.           |
| PE       | 23-DEC-1998;   | 98NZ-0333568.           |
| PR       | 23-JUL-1999;   | 99NZ-0336906.           |
| XX       | (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.              |                         |
| XX       | Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;               |                         |
| XX       | WPI: 2000-452375/39.   |                         |
| XX       | P-PSDB: AAI93750.  |                         |

PT New Perna canaliculus serine protease inhibitor protein exhibiting  
 PT anti-thrombin activity and divalent metal cation binding activity,  
 useful as an anticoagulant agent and as a dietary supplement -  
 PS  
 XX  
 Claim 10: Page 9-10: 44pp; English.  
 CC The present sequence encodes a serine protease inhibitor  
 CC protein. The protein is isolated from the green-lipped mussel  
 CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity  
 CC and divalent metal cation binding activity. The serine protease  
 CC inhibitor protein has a molecular weight of about 55 kilo Daltons.  
 CC The protein, and its fragments, are useful in medicaments, in food,  
 CC as dietary supplements or as bioremediation agents. In the dietary  
 CC supplements, the protein is associated with or bound to at least one  
 CC divalent cation (such as calcium, magnesium or zinc) of dietary  
 CC significance. The proteins or their fragments are also useful as  
 CC anticoagulant agents.  
 CC  
 XX  
 SO Sequence 1491 BP: 428 A; 333 C; 350 G; 379 T; 1 other:

Query Match 100.0%; Score 1490.6; DB 21; Length 1491;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GAYGGGAGCAGTGTACGATGGGAGCAAAAGATGACCCAGACACACGAT 60  
 DB 1 GAYGGGAGCAGTGTACGATGGGAGCAAAAGATGACCCAGACACGAT 60  
 OY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGAT 120  
 DB 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGAT 120  
 OY 121 GAAACCAACCCCTATATGCTGACAGCCCTTCCACCATGTCATGAGCAGATG 180  
 DB 121 GAAACCAACCCCTATATGCTGACAGCCCTTCCACCATGTCATGAGCAGATG 180  
 OY 181 TCACAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 DB 181 TCACAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 OY 241 GAAACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGAT 300  
 DB 241 GAAACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGAT 300  
 OY 301 GGTGATATCTATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 DB 301 GGTGATATCTATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 OY 361 GACCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 DB 361 GACCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 OY 421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 DB 421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
 OY 481 ATTTTACAGGAGTACACACCATGCTGATACCCAGCAGATAGAAATGCTGT 540  
 DB 481 ATTTTACAGGAGTACACACCATGCTGATACCCAGCAGATAGAAATGCTGT 540  
 OY 541 ATTGGATGATGAAAGCTGCGCCAGAAAGAGAGTGTCTACATCAGAGTGA 600  
 DB 541 ATTGGATGATGAAAGCTGCGCCAGAAAGAGAGTGTCTACATCAGAGTGA 600  
 OY 601 GATTAACCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 DB 601 GATTAACCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 OY 661 CTTCATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 DB 661 CTTCATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 OY 721 GTGTCTACCATTTAGAGGATTTAATGATAGATGACCAAAAGATCATCTCAT 780

DB 721 GTGTCTACCATTTAGAGGATTTAATGATAGATGACCAAAAGATCATCTCAT 780  
 OY 781 GTACAGATCTACGCCCAACGGGTGACCATGATGATGATGATGATGATGATGAT 840  
 DB 781 GTACAGATCTACGCCCAACGGGTGACCATGATGATGATGATGATGATGATGAT 840  
 OY 841 GATCCTCATGAAAGATACACAGTGAATGGTGTGATCTGAGATATTCACAGATGAT 900  
 DB 841 GATCCTCATGAAAGATACACAGTGAATGGTGTGATCTGAGATATTCACAGATGAT 900  
 OY 901 CATGGGTTGTCATGAAAGCAGACAGATATTCCTGGATCAATATCTGGGATGAT 960  
 DB 901 CATGGGTTGTCATGAAAGCAGACAGATATTCCTGGATCAATATCTGGGATGAT 960  
 OY 961 GTCCTGGAGCTTCTATTTGCTATTCACCAAGACCATCTTCATTAAGTCCAAAT 1020  
 DB 961 GTCCTGGAGCTTCTATTTGCTATTCACCAAGACCATCTTCATTAAGTCCAAAT 1020  
 OY 1021 GCCTGTTGTCATGACGATGACGACAGCCATCCAGAAATTTGTCACAGACTAAT 1080  
 DB 1021 GCCTGTTGTCATGACGATGACGACAGCCATCCAGAAATTTGTCACAGACTAAT 1080  
 OY 1081 GTTGTACAGCTAATACAGAAATCTACTGTTTACATCACCATGCTCTGTCTAT 1140  
 DB 1081 GTTGTACAGCTAATACAGAAATCTACTGTTTACATCACCATGCTCTGTCTAT 1140  
 OY 1141 TTGCAACAGACCCCTGGAGAGTCAACACATATGACGGTGAATCTCAAGATTT 1200  
 DB 1141 TTGCAACAGACCCCTGGAGAGTCAACACATATGACGGTGAATCTCAAGATTT 1200  
 OY 1201 AGTGAAGACTTGTCAATCATCATGATGATGATGATGATGATGATGATGATGAT 1260  
 DB 1201 AGTGAAGACTTGTCAATCATCATGATGATGATGATGATGATGATGATGATGAT 1260  
 OY 1261 CATGGCTGCTCCTCTTATGAGAGATGATGATGATGATGATGATGATGATGAT 1320  
 DB 1261 CATGGCTGCTCCTCTTATGAGAGATGATGATGATGATGATGATGATGATGAT 1320  
 OY 1321 AGACCTGGAGACCTTGGAGATGATGATGATGATGATGATGATGATGATGAT 1380  
 DB 1321 AGACCTGGAGACCTTGGAGATGATGATGATGATGATGATGATGATGATGAT 1380  
 OY 1381 ACCTTTGAATCTTATGTTGAAGATCTTAACGACGTTCCCTGTGATATTCAG 1440  
 DB 1381 ACCTTTGAATCTTATGTTGAAGATCTTAACGACGTTCCCTGTGATATTCAG 1440  
 OY 1441 GGAATGAGAGTGAAGTGAAGAGGTTGCTGCTGTATATGACGAGGCA 1491  
 DB 1441 GGAATGAGAGTGAAGTGAAGAGGTTGCTGCTGTATATGACGAGGCA 1491

RESULT 2  
 AAA47151  
 ID AAA47151 standard; DNA; 1611 BP.  
 XX  
 AC AAA47151;  
 DE  
 DT 03-OCT-2000 (first entry)  
 XX  
 DE DNA encoding a serine protease inhibitor protein.  
 XX  
 KW Serine protease inhibitor; green-lipped mussel; anti-thrombin;  
 KW divalent metal cation binding activity; dietary supplement;  
 KW anticoagulant; ss.  
 XX  
 OS Perna canaliculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1494  
 FT /\*tag- a  
 FT /product= "serine protease inhibitor"  
 FT polyA\_site 1557..1563

```

XX TT      /tag= b
PN PN      WO200039165-A1.
PD PD      06-JUL-2000.
XX XX
PF PF      23-DEC-1999;    99WO-NZ00227.
PR PR      23-DEC-1999;    98NZ-0333568.
PR PR      23-JUL-1999;    99NZ-0336906.
XX XX
PA PA      (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
PI PI      Scott PD, Dearing SC, Greenwood DR, Newcomb RD;
DR DR      WPI: 2000-452375/39.
PP PP      P-PSDE: AAY93750.
XX XX
PS PS      Claim 11; Page 10-11; 44pp; English.
XX XX
CC CC      The present sequence encodes a serine protease inhibitor
CC CC      (perna canaliculus), and exhibits, inter alia, anti-thrombin activity
CC CC      and divalent metal cation binding activity. The serine protease
CC CC      inhibitor protein has a molecular weight of about 55 kDa Daltons.
CC CC      The protein, and its fragments, are useful in medications, in food,
CC CC      as dietary supplements or as bioremediation agents. In the dietary
CC CC      supplements, the protein is associated with or bound to at least one
CC CC      divalent cation (such as calcium, magnesium or zinc) of dietary
CC CC      significance. The proteins or their fragments are also useful as
XX XX      anticoagulant agents.
S0 S0      Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other;

Query Match          100.0%; Score 1490.6; DB 21; Length 1611;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY   1  GATGGGAGCAGTGTAAACCATTTGGGGCAAAAAGATGACCACATGACGCCACCACGAT 60
DB   1  GATGGGAGCAGTGTAAACCATTTGGGGCAAAAAGATGACCACATGACGCCACCACGAT 60
OY   61  GATCACCATGACGACCATGATGATGATGATGAAGAACAATGCATATGCCAAGTGAATG 120
DB   61  GATCACCATGACGACCATGATGATGATGATGATGAAGAACAATGCATATGCCAAGTGAATG 120
OY   121  GAACCAAAACCTTCATATGGCTATGACACCTTCACCGACCATGTCATGCGAGCATGAGATTG 180
DB   121  GAACCAAAACCTTCATATGGCTATGACACCTTCACCGACCATGTCATGCGAGCATGAGATTG 180
OY   181  TCACAGAAAGGATATGAGAGCTGTTTATTCAGAACACTTCATCTTGTCGATTCACACAGT 240
DB   181  TCACAGAAAGGATATGAGAGCTGTTTATTCAGAACACTTCATCTTGTCGATTCACACAGT 240
OY   241  GAACACCATGACGACCCACCATCATGAGACTTCATGCGACATGCTTGGGACATCTGACGA 300
DB   241  GAACACCATGACGACCCACCATCATGAGACTTCATGCGACATGCTTGGGACATCTGACGA 300
OY   301  GGTTGTATTTCTATTTGGCGAACTGTACATGCTGACGCCAGAAAAACATGCGACCGTGT 360
DB   301  GGTTGTATTTCTATTTGGCGAACTGTACATGCTGACGCCAGAAAAACATGCGACCGTGT 360
OY   361  GACCTCGGTGACCTGTTGACGATGATAGGGCGGTGTTTAATGAAGTTCATCATTAAGT 420
DB   361  GACCTCGGTGACCTGTTGACGATGATAGGGCGGTGTTTAATGAAGTTCATCATTAAGT 420
OY   421  TGTTGGACATTTGTTGATACAGCACCAACACCGAAGCTCTCATTTGACACTCAATGACT 480
DB   421  TGTTGGACATTTGTTGATACAGCACCAACACCGAAGCTCTCATTTGACACTCAATGACT 480

```

|    |      |  |      |
|----|------|--|------|
| OY | 481  | ATTTCACAGGAGATCCACCAGTGTGTTATCCCAGCAGTAGTAATGCCCTGGTGT         | 540  |
| Oy | 481  | TTTTTCAAGGAGATCCACCAGTGTGTTATCCCAGCAGTAGTAATGCCCTGGTGT         | 540  |
| Dd | 481  | ATTTCACAGGAGATCCACCAGTGTGTTATCCCAGCAGTAGTAATGCCCTGGTGT         | 540  |
| OY | 541  | ATTGGTCATGGAAAAGCTCCGCCAAGAACAGACTGCTCTACATCACAGCTAAAGGA       | 600  |
| Oy | 541  | ATTGGTCATGGAAAAGCTCCGCCAAGAACAGACTGCTCTACATCACAGCTAAAGGA       | 600  |
| Dd | 541  | ATTGGTCATGGAAAAGCTCCGCCAAGAACAGACTGCTCTACATCACAGCTAAAGGA       | 600  |
| OY | 601  | GATAAAACTGACCATTTATGCCCCATTGTGCAGTAAAGATCTAATACACCAACCAAAGGCT  | 660  |
| Oy | 601  | GATAAAACTGACCATTTATGCCCCATTGTGCAGTAAAGATCTAATACACCAACCAAAGGCT  | 660  |
| Dd | 601  | GATAAAACTGACCATTTATGCCCCATTGTGCAGTAAAGATCTAATACACCAACCAAAGGCT  | 660  |
| OY | 661  | CTTCATCATCATGATCCACAGSAAACCATCGATTTTCAAMCAAGTGGTTATGTCACCTTGAA | 720  |
| Oy | 661  | CTTCATCATCATGATCCACAGSAAACCATCGATTTTCAAMCAAGTGGTTATGTCACCTTGAA | 720  |
| Dd | 661  | CTTCATCATCATGATCCACAGSAAACCATCGATTTTCAAMCAAGTGGTTATGTCACCTTGAA | 720  |
| OY | 721  | GTCCTCTACCATTTAGAGGGATTATTAATGTAATGATGATGACCCAAGAATCATCTCATAC  | 780  |
| Oy | 721  | GTCCTCTACCATTTAGAGGGATTATTAATGTAATGATGATGACCCAAGAATCATCTCATAC  | 780  |
| Dd | 721  | GTCCTCTACCATTTAGAGGGATTATTAATGTAATGATGATGACCCAAGAATCATCTCATAC  | 780  |
| OY | 781  | GTCACATCTAGCCACGGGACCTGACCACTGATGATGATGATAACCTGGCTCTAAATAT     | 840  |
| Oy | 781  | GTCACATCTAGCCACGGGACCTGACCACTGATGATGATGATAACCTGGCTCTAAATAT     | 840  |
| Dd | 781  | GTCACATCTAGCCACGGGACCTGACCACTGATGATGATGATAACCTGGCTCTAAATAT     | 840  |
| OY | 841  | GATCCCTACAAGATTAATACACATGATGGGTACTAGSAGATATATACATCATATAC       | 900  |
| Oy | 841  | GATCCCTACAAGATTAATACACATGATGGGTACTAGSAGATATATATACATCATATAC     | 900  |
| Dd | 841  | GATCCCTACAAGATTAATACACATGATGGGTACTAGSAGATATATATACATCATATAC     | 900  |
| OY | 901  | CATGGCGTTGTCAATGAAGCCACAGATATTTCTTGATCAATATCTTCGGTATACAGT      | 960  |
| Oy | 901  | CATGGCGTTGTCAATGAAGCCACAGATATTTCTTGATCAATATCTTCGGTATACAGT      | 960  |
| Dd | 901  | CATGGCGTTGTCAATGAAGCCACAGATATTTCTTGATCAATATCTTCGGTATACAGT      | 960  |
| OY | 961  | GTCCTGGGACGTTCTATTGCCATTCACCAAAGACACATCTTCAATAAAGTGCCAAAAT     | 1020 |
| Oy | 961  | GTCCTGGGACGTTCTATTGCCATTCACCAAAGACACATCTTCAATAAAGTGCCAAAAT     | 1020 |
| Dd | 961  | GTCCTGGGACGTTCTATTGCCATTCACCAAAGACACATCTTCAATAAAGTGCCAAAAT     | 1020 |
| OY | 1021 | GCCCTGTTGTATATGAGACGTGACAGACCCATCCAGAAATTTGTCACAGACTAAATGT     | 1080 |
| Oy | 1021 | GCCCTGTTGTATATGAGACGTGACAGACCCATCCAGAAATTTGTCACAGACTAAATGT     | 1080 |
| Dd | 1021 | GCCCTGTTGTATATGAGACGTGACAGACCCATCCAGAAATTTGTCACAGACTAAATGT     | 1080 |
| OY | 1081 | GTTGTACAGCCTAATACAGATCTACAGGTTTACATACACATGCTCTGGTCTTAATACA     | 1140 |
| Oy | 1081 | GTTGTACAGCCTAATACAGATCTACAGGTTTACATACACATGCTCTGGTCTTAATACA     | 1140 |
| Dd | 1081 | GTTGTACAGCCTAATACAGATCTACAGGTTTACATACACATGCTCTGGTCTTAATACA     | 1140 |
| OY | 1141 | TTCGACACACCCCGSGAGATCAACACATAGAGGGTGTATCAAGAGATTAAAGTT         | 1200 |
| Oy | 1141 | TTCGACACACCCCGSGAGATCAACACATAGAGGGTGTATCAAGAGATTAAAGTT         | 1200 |
| Dd | 1141 | TTCGACACACCCCGSGAGATCAACACATAGAGGGTGTATCAAGAGATTAAAGTT         | 1200 |
| OY | 1201 | AGTAGAGACTTGTCAATCATCTCATGTGTGACAGCTCCATGAATGGGAGATATGTCC      | 1260 |
| Oy | 1201 | AGTAGAGACTTGTCAATCATCTCATGTGTGACAGCTCCATGAATGGGAGATATGTCC      | 1260 |
| Dd | 1201 | AGTAGAGACTTGTCAATCATCTCATGTGTGACAGCTCCATGAATGGGAGATATGTCC      | 1260 |
| OY | 1261 | CATGGCTGTACTCTTATAGSCAGATGTACATGTCTCATGATGTAGTGTCTACAGCCCAA    | 1320 |
| Oy | 1261 | CATGGCTGTACTCTTATAGSCAGATGTACATGTCTCATGATGTAGTGTCTACAGCCCAA    | 1320 |
| Dd | 1261 | CATGGCTGTACTCTTATAGSCAGATGTACATGTCTCATGATGTAGTGTCTACAGCCCAA    | 1320 |
| OY | 1321 | AGACCTGTGACCTTGTGAGATTAATAGATGATCCCATGSCATCGTTCATCTACATAGA     | 1380 |
| Oy | 1321 | AGACCTGTGACCTTGTGAGATTAATAGATGATCCCATGSCATCGTTCATCTACATAGA     | 1380 |
| Dd | 1321 | AGACCTGTGACCTTGTGAGATTAATAGATGATCCCATGSCATCGTTCATCTACATAGA     | 1380 |
| OY | 1381 | ACCTTGATCATCTTAATATGTTGAAGATCTTAACGACAGCTCCCTTGTGATATGACGGC    | 1440 |
| Oy | 1381 | ACCTTGATCATCTTAATATGTTGAAGATCTTAACGACAGCTCCCTTGTGATATGACGGC    | 1440 |
| Dd | 1381 | ACCTTGATCATCTTAATATGTTGAAGATCTTAACGACAGCTCCCTTGTGATATGACGGC    | 1440 |
| OY | 1441 | GAGCATGAGGTGAGAGTGAAGAGGTGCTGTGTTATAGACGGGCA 1491              |      |
| Oy | 1441 | GAGCATGAGGTGAGAGTGAAGAGGTGCTGTGTTATAGACGGGCA 1491              |      |
| Dd | 1441 | GAGCATGAGGTGAGAGTGAAGAGGTGCTGTGTTATAGACGGGCA 1491              |      |

[illegible]

|          |   |   |     |
|----------|---|---|-----|
| Db       | 379   | GATGGTCACATGATATATGATGATGGTGATATATGATGATATGATATGATATGATATGAT    | 438 |
| Oy       | 181   | TCACAGAGAGCGTCATGACAGCTGTTATCTAGAACTTCATCTTGTCGATTCACACAAGT     | 240 |
| Db       | 439   | GATGGTCATGATGATGATGGTGTTGATGATGATGATGATGATGATGATGATGATGAT       | 498 |
| Oy       | 241   | GAGACCATGACGACACCACCATCATGAGCTTCATCTGACACATCTTGTCGACATTCAGCA    | 300 |
| Db       | 499   | GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 558 |
| Oy       | 301   | GTTTGTGATTCATTATGGGCAACTGTACAAATGCTCCACCAGAAAAACATGCTGACCCCTGGT | 360 |
| Db       | 559   | AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 618 |
| Oy       | 361   | GACCTGCGTACCTGCTGGTGACGATGATAGAGGCGGTGATTAATGAAGTTCATCATTTATGCT | 420 |
| Db       | 619   | GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT     | 678 |
| Oy       | 421   | TG 422  |     |
| Db       | 679   | GG 680  |     |
| RESULT 4 |   |   |     |
| ABAS0512 |   |   |     |
| ID       | ABAS0512 standard; DNA; 583 BP.   |   |     |
| XX       |   |   |     |
| MC       | ABAS0512:   |   |     |
| XX       |   |   |     |
| DT       | 01-FEB-2002 (first entry)   |   |     |
| XX       |   |   |     |
| DE       | Human breast cell single exon nucleic acid probe #9207.   |   |     |
| XX       |   |   |     |
| KW       | Human; microarray; single exon probe; gene expression; breast; disease; cancer; ss.   |   |     |
| XX       |   |   |     |
| OS       | Homo sapiens.   |   |     |
| XX       |   |   |     |
| PN       | W0200157271-A2.   |   |     |
| PD       | 09-AUG-2001.  |   |     |
| XX       |   |   |     |
| PF       | 30-JAN-2001; 2001WO-US00662.  |   |     |
| XX       |   |   |     |
| PR       | 04-FEB-2000; 2000US-0180312.  |   |     |
| PR       | 26-MAY-2000; 2000US-0207456.  |   |     |
| PR       | 30-JUN-2000; 2000US-0608408.  |   |     |
| PR       | 03-AUG-2000; 2000US-0632366.  |   |     |
| PR       | 21-SEP-2000; 2000US-0234687.  |   |     |
| PR       | 27-SEP-2000; 2000US-0236359.  |   |     |
| PR       | 04-OCT-2000; 2000GB-0024263.  |   |     |
| XX       |   |   |     |
| PA       | (MOLE-) MOLECULAR DYNAMICS INC.   |   |     |
| XX       |   |   |     |
| PI       | Penn SG, Hanzel DK, Chen W, Rank DR;  |   |     |
| XX       |   |   |     |
| DR       | WPI; 2001-496933/54.  |   |     |
| XX       |   |   |     |
| PT       | New spatially-addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes   |   |     |
| XX       |   |   |     |
| XX       | Claim 4; SEQ ID NO 9207; 327bp + sequence listing; English.   |   |     |
| XX       |   |   |     |
| XX       | The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labeled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene |   |     |





PD 09-AUG-2001.  
 XX  
 XX  
 PE 30-JAN-2001; 2001WO-US00666.  
 XX  
 XX 04-FEB-2000; 2000US-0180312.  
 FR 26-MAY-2000; 2000US-0207456.  
 FR 30-JUN-2000; 2000US-0608408.  
 FR 03-AUG-2000; 2000US-0632366.  
 FR 21-SEP-2000; 2000US-0234587.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 P1 Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR MPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 4; SEQ ID No 13918; 530pp: English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart. The  
 CC present sequence is one such probe. The probes may be used for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from the human heart via microarrays. By measuring gene expression, the  
 CC probes are useful for predicting, diagnosing, grading, staging,  
 CC monitoring and prognosing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
 XX  
 SQ Sequence 583 BP: 172 A; 38 C; 212 G; 161 T; 0 other;  
 XX

|    | Query Match  | Similarity   | 3.5%;  | Score 52.6;        | DB 22;    | Length 583; |
|----|--------------|--|--------|--------------------|-----------|-------------|
|    | Best Local   | Similarity   | 47.0%; | Pred. No. 7.8e-05; |           |             |
|    | Matches 194; | Conservative   | 1;     | Mismatches 215;    | Indels 3; | Gaps 1.     |
| QY | 1            | GATGGGAGCAGTGTAAACGATGGCGCAGAACAAAGTAGACCCACCATGACACCCACGAT      | 60     |                    |           |             |
| Db | 41           | GATGGGAGCAGTGTAAACGATGGCGCAGAACAAAGTAGACCCACCATGACACCCACGAT      | 100    |                    |           |             |
| QY | 61           | GATCCCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT   | 120    |                    |           |             |
| Db | 101          | GAT     | 160    |                    |           |             |
| QY | 121          | GATCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT   | 180    |                    |           |             |
| Db | 161          | GATCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT   | 220    |                    |           |             |
| QY | 181          | TCACAGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT | 240    |                    |           |             |
| Db | 221          | AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT      | 280    |                    |           |             |
| QY | 241          | GAT     | 300    |                    |           |             |
| Db | 281          | GAT     | 337    |                    |           |             |
| QY | 301          | GAT     | 360    |                    |           |             |
| Db | 338          | GAT     | 397    |                    |           |             |
| QY | 361          | GAT     | 413    |                    |           |             |
| Db | 398          | GAT     | 450    |                    |           |             |

AKI6834  
ID AKI6834 standard DNA; 583 BP.  
XX AC AKI6834:  
XX XX  
DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe SEQ ID NO: 16925.  
XX KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
RW epilepsy; cancer; ss.  
XX OS Homo sapiens.  
XX PN M0200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 03-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI: 2001-483446/52.  
XX DR  
XX PT Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX PS Example 4; SEQ ID NO: 16925; 650bp + Sequence Listing; English.  
XX SS  
XX CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX CC  
XX CC  
SQ Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;

Query Match 3.5%; Score 52.6; DB 22; Length 583;  
Best Local Similarity 47.0%; Pred. No. 7.8e-05;  
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1.

1  
GAYGGGAGGAGTGTACGATGGCGGAACAAGATGCACCACATGACGCCACCAAGAT 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
41 GATGGGAATATGTGTATGATGTGTATGATGACACTGATGATGATGATGATGATGG 100  
61 GATCACCATACGACCATGATGATGATGATGAAACAATGCACTTGCCCACTGGAATTG 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
101 GATGATGATACAGTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 160  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
121 GAACCAACCCCTCATTTGGCTTAGCAGCCTTCACCAACCATGTCATGGCAGCATAGA 180  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
161 GGGGAATATGATGACAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 220  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
181 TCACAGAAAGGATGATGAGACTGTTATCTAGAACTTCATCTTGTGGATTCACACAGT 240  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
221 AGTATGATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 280  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
241 GAAGACCATACGACCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
281 GAT 337

```
Oy 301 GGTGTGATTCATTATGCGACATGATACATGTCACCCAGAAAAACATGCTGACCCCTGG 360
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 338 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 361 GACCTCGTACCTGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATG 413
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450

RESULT 8
AAK42604
ID AAK42604 standard; DNA; 583 BP.
XX
AC AAK42604;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 17161.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX
PS Example 4; SEQ ID NO: 17161; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;
XX

Query Match 3.5%; Score 52.6; DB 22; Length 583;
Best Local Similarity 47.0%; Pred. No. 7.8e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
```

```
Db 161 GGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 220
Oy 181 TCACAGACGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 221 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 241 GAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 301 GGTGTGATTCATTATGCGAACATGCTACATGTCACCCAGAAAAACATGCTGACCCCTGG 360
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 338 GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 361 GACCTCGTACCTGCTGACATGATGATGATGATGATGATGATGATGATGATGATGATG 413
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450

RESULT 9
AAI23356
ID AAI23356 standard; DNA; 583 BP.
XX
AC AAI23356;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #13289 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 25; SEQ ID NO 13289; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp://ipb.int/pub/published_pcl_sequences.
XX
XX Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;
XX

Query Match 3.5%; Score 52.6; DB 22; Length 583;
```

Best Local Similarity 47.0%; Pred. No. 7.8e-05;  
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;

```
OY 1 GAYGGGAGCAGTGTAAACGATGGGAGACAAAGATGACACCATGACACACACAGAT 60
    ||:|||||
DB 41 GATGGGATGATGGTGTATGGTGTATGATGACAGTATGATGATGATGATGATGG 100
OY 61 GATCAGCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATG 120
    ||:|||||
DB 101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 160
OY 121 GACCAAAACCTCATATGCTAGACGCTTACCACCATGTCATGACATGAGATTG 180
    ||:|||||
DB 161 GGGGAAGATGATGACAGTATGATGATGATGATGATGATGATGATGATGATG 220
OY 181 TCACAGAAAGGTCATGAGCTGTTTATCTAGAACTTCATCTGCGATTGACACAAGT 240
    ||:|||||
DB 221 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
OY 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATG 300
    ||:|||||
DB 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 337
OY 301 GGTGTGATTCATTTGGGCACTGTACATGCTCACCAGAAAACATGCTGACCTGGT 360
    ||:|||||
DB 338 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
OY 361 GACCTCGGTGACCTGTTGACGATGATGATGATGATGATGATGATGATGATG 413
    ||:|||||
DB 398 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
```

RESULT 10

AA148680 standard; DNA; 583 BP.

AA148680;

17-OCT-2001 (first entry)

Probe #17366 used to measure gene expression in human placenta sample.

Probe; microarray; human; placenta; antenatal diagnosis; ss;

genetic disorder; ss.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632386.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488897/53.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human placenta -

Claim 25; SEQ ID NO 17366; 654bp; English.

The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

Sequence 583 BP; 172 A; 38 C; 212 G; 161 T; 0 other;

Query Match 3.5%; Score 52.6; DB 22; Length 583;

Best Local Similarity 47.0%; Pred. No. 7.8e-05;  
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;

```
OY 1 GAYGGGAGCAGTGTAAACGATGGGAGACAAAGATGACACCATGACACACACAGAT 60
    ||:|||||
DB 41 GATGGGATGATGGTGTATGGTGTATGATGACAGTATGATGATGATGATGATGG 100
OY 61 GATCAGCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATG 120
    ||:|||||
DB 101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 160
OY 121 GACCAAAACCTCATATGCTAGACGCTTACCACCATGTCATGACATGAGATTG 180
    ||:|||||
DB 161 GGGGAAGATGATGACAGTATGATGATGATGATGATGATGATGATGATGATG 220
OY 181 TCACAGAAAGGTCATGAGCTGTTTATCTAGAACTTCATCTGCGATTGACACAAGT 240
    ||:|||||
DB 221 AGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
OY 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATG 300
    ||:|||||
DB 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 337
OY 301 GGTGTGATTCATTTGGGCACTGTACATGCTCACCAGAAAACATGCTGACCTGGT 360
    ||:|||||
DB 338 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
OY 361 GACCTCGGTGACCTGTTGACGATGATGATGATGATGATGATGATGATGATG 413
    ||:|||||
DB 398 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
```

RESULT 11

AA108992 standard; DNA; 583 BP.

AA108992;

09-OCT-2001 (first entry)

Probe #8983 used to measure gene expression in human breast sample.

Probe; human; breast disease; breast cancer; development disorder; ss;

Inflammatory disease; proliferative breast disease; non-carcinoma tumour.

Homo sapiens.

WO200157270-A2.

09-AUG-2001.

29-JAN-2001; 2001WO-US00661.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632386.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488897/53.







Mon Dec 2 12:19:29 2002

us-09-868-760-6.rng

Page 12

[illegible]

Search completed: November 29, 2002, 07:40:20  
Job time : 289 secs

•



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 06:39:33 ; Search time 50 Seconds  
(Without alignments)  
11484.142 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gagggggagcagtgtaacga.....gctgtgtatagtagcagggca 1491

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 341543 segs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCNT\_NEM\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCNTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match % | Length | ID                  | Description       |
|------------|-------|---------------|--------|---------------------|-------------------|
| 1          | 52.6  | 3.5           | 583    | US-09-864-761-20772 | Sequence 20772, A |
| 2          | 52.6  | 3.5           | 1959   | US-09-864-761-4012  | Sequence 4012, Ap |
| 3          | 49.2  | 3.3           | 327    | US-09-864-761-28059 | Sequence 28059, A |
| 4          | 46.2  | 3.1           | 456    | US-09-864-761-11468 | Sequence 11468, A |
| 5          | 44.2  | 3.0           | 574    | US-09-864-761-228   | Sequence 228, App |
| 6          | 44.2  | 3.0           | 659    | US-09-864-761-17051 | Sequence 17051, A |
| 7          | 43.4  | 2.9           | 470    | US-09-864-761-3121  | Sequence 3121, Ap |
| 8          | 43.4  | 2.9           | 522    | US-09-864-761-19900 | Sequence 19900, A |
| 9          | 43.2  | 2.9           | 537    | US-09-864-761-8330  | Sequence 8330, Ap |
| 10         | 43.2  | 2.9           | 1075   | US-09-864-761-19241 | Sequence 19241, A |
| 11         | 43.2  | 2.9           | 1403   | US-09-864-761-2513  | Sequence 2513, Ap |
| 12         | 42.6  | 2.9           | 1959   | US-09-864-761-4012  | Sequence 4012, Ap |
| 13         | 42.4  | 2.8           | 703    | US-09-910-943-302   | Sequence 302, App |
| 14         | 42.4  | 2.8           | 766    | US-09-864-761-19608 | Sequence 19608, A |
| 15         | 42.4  | 2.8           | 1944   | US-09-864-761-2825  | Sequence 2825, Ap |
| 16         | 41.4  | 2.8           | 276    | US-09-864-761-23120 | Sequence 23120, A |
| 17         | 41    | 2.7           | 240    | US-09-864-761-23948 | Sequence 23948, A |
| 18         | 41    | 2.7           | 554    | US-09-864-761-7217  | Sequence 7217, Ap |
| 19         | 40.6  | 2.7           | 390    | US-09-790-399-7     | Sequence 7, Appl1 |

|    |      |     |       |    |                     |                    |
|----|------|-----|-------|----|---------------------|--------------------|
| 20 | 40.4 | 2.7 | 672   | 9  | US-09-938-842A-2202 | Sequence 2202, Ap  |
| 21 | 40.2 | 2.7 | 366   | 10 | US-09-974-300-6355  | Sequence 6355, Ap  |
| 22 | 39.8 | 2.7 | 2403  | 10 | US-09-815-242-9287  | Sequence 9287, Ap  |
| 23 | 39.6 | 2.7 | 2069  | 10 | US-09-842-552-17    | Sequence 17, Appl1 |
| 24 | 38.8 | 2.6 | 439   | 10 | US-09-864-761-20117 | Sequence 20117, A  |
| 25 | 38.8 | 2.6 | 441   | 10 | US-09-864-761-3403  | Sequence 3403, Ap  |
| 26 | 38.8 | 2.6 | 444   | 10 | US-09-864-761-17518 | Sequence 27518, A  |
| 27 | 38.8 | 2.6 | 478   | 10 | US-09-728-446-1415  | Sequence 27518, A  |
| 28 | 38.8 | 2.6 | 491   | 10 | US-09-864-761-10879 | Sequence 10879, A  |
| 29 | 38.8 | 2.6 | 517   | 12 | US-10-033-528-313   | Sequence 313, App  |
| 30 | 38.8 | 2.6 | 517   | 10 | US-09-920-300A-313  | Sequence 313, App  |
| 31 | 38.8 | 2.6 | 20029 | 10 | US-09-764-847-1624  | Sequence 1624, A   |
| 32 | 38.6 | 2.6 | 331   | 10 | US-09-864-761-17053 | Sequence 17053, A  |
| 33 | 38.6 | 2.6 | 446   | 10 | US-09-864-761-20699 | Sequence 20699, A  |
| 34 | 38.6 | 2.6 | 465   | 10 | US-09-864-761-230   | Sequence 230, App  |
| 35 | 38.4 | 2.6 | 2581  | 10 | US-09-864-761-230   | Sequence 66, Appl  |
| 36 | 38.2 | 2.6 | 423   | 9  | US-09-938-842A-592  | Sequence 592, App  |
| 37 | 38.2 | 2.6 | 565   | 10 | US-09-864-761-8619  | Sequence 8619, Ap  |
| 38 | 38.2 | 2.6 | 583   | 10 | US-09-864-761-20772 | Sequence 20772, A  |
| 39 | 37.8 | 2.5 | 131   | 10 | US-09-864-761-18475 | Sequence 18475, A  |
| 40 | 37.8 | 2.5 | 258   | 12 | US-10-028-247-1     | Sequence 1, Appl1  |
| 41 | 37.8 | 2.5 | 491   | 10 | US-09-864-761-1720  | Sequence 1720, Ap  |
| 42 | 37.4 | 2.5 | 420   | 10 | US-09-864-761-23266 | Sequence 23266, A  |
| 43 | 37.4 | 2.5 | 451   | 10 | US-09-864-761-6554  | Sequence 6554, Ap  |
| 44 | 37.4 | 2.5 | 473   | 10 | US-09-864-761-4184  | Sequence 4184, Ap  |
| 45 | 37.4 | 2.5 | 1411  | 10 | US-09-191-724-5     | Sequence 5, Appl1  |

#### ALIGNMENTS

RESULT 1  
US-09-864-761-20772  
Sequence 20772, Appl1 Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30

```
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 20772
;; LENGTH: 583
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL022334.1
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
;; OTHER INFORMATION: EST_HUMAN HIT: AW923395.1, EVALU8 9.50e-02
US-09-864-761-20772
```

```
Query Match 3.5%; Score 52.6; DB 10; Length 583;
Best Local Similarity 47.0%; Pred. No. 7.3e-06;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
```

```
QY 1 GAYGGGAGCAGTGTAAAGATGAGCAACAAAGATGACCAACATGACAGCACCACGAT 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 41 GATGGGAATGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GATCACCATGACGACATGATGATGATGATGATGATGATGATGATGATGATGATG 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 101 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 160
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GAACCAACCTCATATGCTAGCAGCCTTACCACCATGCTGATGATGATGATGATG 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 161 GCGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 220
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 TCACGAAGGCGATGAGCTGTTATCTAGAACTTATCTTGTGGATTCACACAGT 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 221 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GAAGACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATG 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 281 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GGTGTGATTCATTTGGGCACTGTACATGCTCACCAGAAACATGCTGACCTGCT 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 338 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 GACCTCGGTGACCTGTTGACGATGATGATGATGATGATGATGATGATGATGAT 413
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 398 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 450
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

## RESULT 2

```
US-09-864-761-4012
;; Sequence 4012, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
```

```
;; FILE REFERENCE: Aeomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263,6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 4012
;; LENGTH: 1959
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL022334.1
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012
```

```
Query Match 3.5%; Score 52.6; DB 10; Length 1959;
Best Local Similarity 47.0%; Pred. No. 1.6e-05;
Matches 194; Conservative 1; Mismatches 215; Indels 3; Gaps 1;
```

```
QY 1 GAYGGGAGCAGTGTAAAGATGAGCAACAAAGATGACCAACATGACAGCACCACGAT 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 546 GATGGGAATGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 605
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 606 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 665
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GAACCAACCTCATATGCTAGCAGCCTTACCACCATGCTGATGATGATGATGATG 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db      666 GGGGAAGATGATGACAGCATGTGTGCATGATGACAGATAAGCAATGATGGGATGATGAC    725
Oy      181 TCACAACAGCGTCATGAGCTGTTTATTCTTGAACTTCATCCTTGATTAACACAACT   240
Db      726 AAGAATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG   785
Oy      241 GAAGACCATGACGACACCACTCATGACTTCATCTCATCTCCATCTGTCGACATCTGCA    300
Db      786 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG   842
Oy      301 GGTCGTGATTCATTTATGGCACTGTACATATCTCACCCAGAAAAAATCTGACCTCGT    360
Db      843 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG   902
Oy      361 GACCTCGTGACCTCGTTGACGATGATAGGCGCTGTTAATGAATTCATCA    413
Db      903 AATGGATGATGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGA    955

RESULT 3
US-09-864-761-28059/c
; Sequence 28059, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864, 761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28059
LENGTH: 327
```

[illegible]

```

PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 11468
LENGTH: 456
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005668.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 0.55
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5

```

[illegible]

```

PRIOR APPLICATION NUMBER: US 09/653,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263, 6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 228
LENGTH: 574

```

| Query Match   | Best Local Similarity | Matches   | Conservative    | Score       | Pred. | No. 0.0024 | Mismatches | Indels | Gaps |
|---|-----------------------|---|-----------------|-------------|-------|------------|------------|--------|------|
| ORGANISM: Homo sapiens                                    |                       |   |                 |             |       |            |            |        |      |
| FEATURE:  |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: MAP TO A1079301.14                     |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN HEPL100, SIGNAL = 7.8     |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.2 |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8.2       |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.4    |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.3       |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.6 |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.2        |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.8 |                       |   |                 |             |       |            |            |        |      |
| OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7         |                       |   |                 |             |       |            |            |        |      |
| US-09-864-761-228   |                       |   |                 |             |       |            |            |        |      |
| Query Match   | 3.0%                  | Score 44.2:   | DB 10:          | Length 574: |       |            |            |        |      |
| Best Local Similarity                                     | 47.9%                 | Pred. No. 0.0024:   |                 |             |       |            |            |        |      |
| Matches 127:  | Conservative          | 0:  | Mismatches 138: | Indels      | 0:    | Gaps       | 0:         |        |      |
| Oy  | 15                    | TAAACGATGGGAGAAACAAAGTACACCCATGACGACACAGATGATCACCATGACGA      | 74              |             |       |            |            |        |      |
| Db  | 116                   | TCACATGATGCATCCATCACCACCCATCACCAGAGACACACCATCACCATGACGATGACGA | 175             |             |       |            |            |        |      |
| Oy  | 75                    | CCATGTGATGTATGATGAACAATGACATGATGCCAGTGGGAATGGAACCAAACCTGCA    | 134             |             |       |            |            |        |      |
| Db  | 176                   | CCACCATCACCATGACGACACACACACACATGACATTCACCATTTATCAGCAACA       | 235             |             |       |            |            |        |      |
| Oy  | 135                   | TATGGCTAGACACCTTTACACGACATGTCCATGCGACAGATAGATTTGTACAGAAAGGTCA | 194             |             |       |            |            |        |      |
| Db  | 236                   | CTATGCAACCATCTCTATCACCATCACCATGACACCATTCACACGACGACACACCA      | 295             |             |       |            |            |        |      |
| Oy  | 195                   | TGGAGCTGTTATCTAGAACCTTCCTGTGCGATTCAACAAGTGAAGACCATGACGA       | 254             |             |       |            |            |        |      |
| Db  | 296                   | TCACCATGAGACACACACACATTCACATGACATGATTCACCATTTATCACCATCACTATGA | 355             |             |       |            |            |        |      |
| Oy  | 255                   | CCACCATCATGAGCTTCATCTGAC                                      | 279             |             |       |            |            |        |      |



PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 3121  
 LENGTH: 470  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AL049749.1  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2  
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.7  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9  
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.4  
 US-09-864-761-3121  
 Query Match  
 Best Local Similarity 55.7%; Score 43.4; DB 10; Length 470;  
 Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
 QY 17 ACAGTGGCAGCAAGATGACGACCATGACGACCCAGCATGATGATGACGACG 76  
 Db 160 ACCATCATGACCCAGCCATCATCATCATCAACACGACCATCATGACCATGAC 101  
 QY 77 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 136  
 Db 100 ACCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 41  
 QY 137 TGGCTGACGAGCTTACGACCATGATGATGATGATGATGATGATGATGATGAT 165  
 Db 40 ACTACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12  
 RESULT 8  
 US-09-864-761-19900/c  
 Sequence 19900 Application US/09864761  
 Patient No US2002048763A1  
 GENERAL INFORMATION:  
 APPLICANT: Penn, Sharon G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 FILE REFERENCE: Aecm1ca-X-1  
 CURRENT FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US/09/864,761  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: US 09/632,366  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: GB 24263.6  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00665

```

PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,667
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
SEQ ID NO: 19900
LENGTH: 522
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL049749.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
OTHER INFORMATION: EST_HUMAN HIT: NA176664.1, EVALUATE 4.90e-01
US-09-864-761-19900

Query Match      2.9%; Score 43.4; DB 10; Length 522;
Best Local Similarity 55.7%; Pred.No. 0.004;
Matches 83; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 17 ACGATGGCAGAAACAAGAATGATGCACCACCATGTGACGCCATGACGACC 76
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 ACCATATGATCAACCCACCCATCATCATTCACACGACCATGACCATCACACC 362
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 77 ATGATGATGATGATGAACATGACACTATGCCCATGTGTGAATAATGAAACCAACCTTCATTA 136
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 361 ACCTTGATGATGATCACTGCCACCTGCATCATGTGCTATTATTAACACCCGCGATCACCATCAATC 302
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 137 TGGCTAGCAGCCCTTCCACCCACCATGTGCAT 165
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 ACTAACCCATCATCATCACCATCATCATCAT 273
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-864-761-8390
Sequence 8390, Application us/09864761
Patient No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon R.
APPLICANT: Bank, David G.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
```

```

PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
SEQ ID NO 8390
LENGTH: 337
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC012153.10
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
US-09-864-761-8390

Query Match      2.9%  Score 43.2;  DB 10;  Length 537;
Best Local Similarity 47.7%;  Pred. No. 0.0046;
Matches 126;  Conservative 0;  Mismatches 138;  Indels 0;  Gaps 0;

QY 17  ACGATGGCGAAGAAAGATGAGACCAATGACGACGACACGATGATGACGACGACG 76
DB 6  ACCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 65
QY 77  ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 136
DB 66  ATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 125
QY 137  TGGCTACACCTTCACCAACCATGTCATGACGACGACGATGATGATGATGATGATG 196
DB 126  GTCACCAACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 185
QY 197  GAGCTGTTATCTAAGAACTTCATCTGTGATGATGATGATGATGATGATGATGATG 256
DB 186  ATCAACCAACCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 245
QY 257  ACCATCATGACCTTCATGTCACA 280
```

```

DB 246  ACCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 269
|||||  |||  |||  |||  |||
RESULT 10
US-09-864-761-19241/C
Sequence 19241, Application US/09864761
Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecolica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR APPLICATION NUMBER: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR APPLICATION NUMBER: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR APPLICATION NUMBER: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
SEQ ID NO 19241
LENGTH: 1075
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL078472.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
OTHER INFORMATION: EXPRESSED IN HL100, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
OTHER INFORMATION: EST_HUMAN HIT: AV739739.1, EVALUE 1.00e+00
```





```

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 4012
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022334.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012

Query Match
Best Local Similarity 5.9%; Score 42.6; DB 10; Length 1959;
Matches 81; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

OY 8 AGCAGCTTACCATGGCGAAGAAAGATGACCCATGACGACACCATGATCACC 67
DB 533 ATATTATACCATTTATCATCTACCCCTTACCCACCATCATCATCATCATCACC 474
OY 68 ATGACGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 127
DB 473 ATCACCACCATATCATCATCATTTTACCATCATCATCATCATCATCATCATC 414
OY 128 ACCCTCATATGGCTAGCAGCCTTCA 152
DB 413 AGCGTCATCATCATGAGCCACATCA 389

RESULT 13
US-09-910-943-302
; Sequence 302, Application US/09910943
; Patent No. US20020081610A1
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Bryanlou, Ali
; APPLICANT: Altman, Curtis
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
; FILE REFERENCE: 7529/16140US1
```

```

; CURRENT APPLICATION NUMBER: US/09/910,943
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 742
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 302
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)..(703)
; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-302

Query Match
Best Local Similarity 2.8%; Score 42.4; DB 10; Length 703;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

OY 32 AAGATGACCACTATGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 91
DB 402 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 461
OY 92 AATCATGCTACTATGCCCGAGTGAATGGAACCAACCC 131
DB 462 AATAATGATTGTCTGCTGATTAATTTGTCTGAAGCCC 501

RESULT 14
US-09-864-761-19608/C
; Sequence 19608, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: CB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
```

```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19608
; LENGTH: 766
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P17164, EVALU 4.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AM844901.1, EVALU 2.90e+00
US-09-864-761-19608

Query Match          2.8%; Score 42.4; DB 10; Length 766;
Best Local Similarity 46.9%; Pred. No. 0.01;
Matches 199; Conservative 0; Mismatches 221; Indels 4; Gaps 2;

QY 19 GATGGGCAAGCAAGATGACACCATGACGACCCACGATGATCAACATGACGACCAT 78
DB 621 GAGCATGATATATATATGATGACGATGATGATGATGATGATGATGATGATGATG 562
QY 79 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
DB 561 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 502
QY 139 --GCTAGACGCTTCCACCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 196
DB 501 GTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 442
QY 197 GAGCTGTTATCTGAACTTCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256
DB 441 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 382
QY 257 ACCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 314
DB 381 ACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 322
QY 315 TGGGCAACTGTACATGCTCCACCAAGAAACATGCTGCTGCTGCTGCTGCTGCTG 374
DB 321 TGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 262
QY 375 GATTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 434
DB 261 TGTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 202
QY 435 TGGT 438
DB 201 TGGT 198

RESULT 15
US-09-864-761-2825/C
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```

; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1825
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825

Query Match          2.8%; Score 42.4; DB 10; Length 1944;
Best Local Similarity 46.9%; Pred. No. 0.018;
Matches 199; Conservative 0; Mismatches 221; Indels 4; Gaps 2;

QY 19 GATGGGCAAGCAAGATGACACCATGACGACCCACGATGATCAACATGACGACCAT 78
DB 858 GAGCATGATATATATATGATGACGATGATGATGATGATGATGATGATGATGATG 799
QY 79 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 138
DB 798 GAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 739
QY 139 --GCTAGACGCTTCCACCAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 196
DB 738 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 679
QY 197 GAGCTGTTATCTGAACTTCACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256
DB 678 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 619
QY 257 ACCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 314
```

```

Db 618 ACAATGATGATGATTATGATTAATGATGATGAGAGATGAGAAATGAGATGATATGG 559
QY 315 TGGCGAACTGTACATGCTCACACCCAGAAAAACATGCTGACCTGTGACCTCGGTGACCT 374
Db 558 TGACAAATGGTAAATGATGGTGAATGAGAGACGATGCGGATGATGATGAAGATGATGA 499
QY 375 GGTGACGATGATGAGGCGGTGTTAATGAGTTCATCATTTATGCTGGTTGACATGA 434
Db 498 TGTGGCGATGAGAGATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439
QY 435 TGGT 438
Db 438 TGGT 435

```

Search completed: November 29, 2002, 07:43:03  
 Job time : 70 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 10:09:03 ; Search time 2876 Seconds

(without alignments)  
15087.720 Million cell updates/sec

Title: US-09-868-760-6

Sequence: 1 gcyggggagcagtgtaacga.....gctgtgtatagagcggca 1491

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*\*  
1: gb\_ha:\*\*  
2: gb\_hg:\*\*  
3: gb\_in:\*\*  
4: gb\_om:\*\*  
5: gb\_ov:\*\*  
6: gb\_pa:\*\*  
7: gb\_ph:\*\*  
8: gb\_pl:\*\*  
9: gb\_pr:\*\*  
10: gb\_ro:\*\*  
11: gb\_sts:\*\*  
12: gb\_sy:\*\*  
13: gb\_un:\*\*  
14: gb\_vl:\*\*  
15: em\_ba:\*\*  
16: em\_fun:\*\*  
17: em\_hum:\*\*  
18: em\_in:\*\*  
19: em\_mu:\*\*  
20: em\_om:\*\*  
21: em\_or:\*\*  
22: em\_ov:\*\*  
23: em\_pat:\*\*  
24: em\_ph:\*\*  
25: em\_pl:\*\*  
26: em\_ro:\*\*  
27: em\_sts:\*\*  
28: em\_un:\*\*  
29: em\_vl:\*\*  
30: em\_htg\_hum:\*\*  
31: em\_htg\_inv:\*\*  
32: em\_htg\_mus:\*\*  
33: em\_htg\_mus:\*\*  
34: em\_htg\_pln:\*\*  
35: em\_htg\_rtd:\*\*  
36: em\_htg\_mam:\*\*  
37: em\_htg\_vrt:\*\*  
38: em\_sy:\*\*  
39: em\_htgo\_hum:\*\*  
40: em\_htgo\_mus:\*\*  
41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID         | Description         |
|------------|-------|-------------|--------|----|------------|---------------------|
| 1          | 1422  | 95.4        | 1700   | 3  | AF273766   | AF273766 Perna can  |
| 2          | 26    | 1.7         | 10210  | 3  | AC020131   | AC020131 Drosophila |
| 3          | 26    | 1.7         | 17735  | 3  | AC005714   | AC005714 Drosophila |
| 4          | 26    | 1.7         | 307363 | 3  | AE003457   | AE003457 Drosophila |
| 5          | 25    | 1.7         | 65259  | 2  | AC108920   | AC108920 Mus muscu  |
| 6          | 25    | 1.7         | 67759  | 2  | AC019858   | AC019858 Drosophila |
| 7          | 25    | 1.7         | 192588 | 3  | AC010057   | AC010057 Drosophila |
| 8          | 25    | 1.7         | 207018 | 2  | AE003469   | AE003469 Drosophila |
| 9          | 25    | 1.7         | 302527 | 3  | AE003469   | AE003469 Drosophila |
| 10         | 24    | 1.6         | 476    | 8  | AF231797   | AF231797 Pinnus tae |
| 11         | 24    | 1.6         | 508    | 8  | AF172408   | AF172408 Pinnus tae |
| 12         | 24    | 1.6         | 179289 | 10 | AL590633   | AL590633 Mouse DNA  |
| 13         | 24    | 1.6         | 209173 | 2  | AL845279   | AL845279 Mus muscu  |
| 14         | 22    | 1.5         | 1817   | 1  | BRUBCSP    | BRUBCSP             |
| 15         | 22    | 1.5         | 9960   | 1  | AE009521   | AE009521 Brucella   |
| 16         | 22    | 1.5         | 56956  | 2  | AC117081   | AC117081 Dictyoste  |
| 17         | 22    | 1.5         | 135691 | 2  | AC095233   | AC095233 Rattus no  |
| 18         | 22    | 1.5         | 156312 | 2  | AC121174   | AC121174 Rattus no  |
| 19         | 22    | 1.5         | 165864 | 2  | AC123157   | AC123157 Rattus no  |
| 20         | 22    | 1.5         | 168909 | 2  | AC017470   | AC017470 Drosophila |
| 21         | 22    | 1.5         | 169289 | 3  | AC007145   | AC007145 Drosophila |
| 22         | 22    | 1.5         | 265605 | 3  | AE003562   | AE003562 Drosophila |
| 23         | 22    | 1.4         | 24     | 6  | A20717     | A20717 Oligonucleo  |
| 24         | 22    | 1.4         | 33     | 6  | AL9463     | AL9463 Oligonucleo  |
| 25         | 22    | 1.4         | 23     | 3  | AF286896   | AF286896 Plasmodiu  |
| 26         | 22    | 1.4         | 706    | 3  | AF223572   | AF223572 Hirudo ni  |
| 27         | 22    | 1.4         | 846    | 9  | AB044395S4 | AB044395 Hirudo ni  |
| 28         | 22    | 1.4         | 1587   | 8  | AY084494   | AY084494 Arabidops  |
| 29         | 22    | 1.4         | 1965   | 3  | DM073490   | DM073490 Arabidops  |
| 30         | 22    | 1.4         | 2108   | 3  | AF272790   | AF272790 Hirudo ni  |
| 31         | 22    | 1.4         | 2922   | 3  | AF071417   | AF071417 Drosophila |
| 32         | 22    | 1.4         | 2980   | 3  | AY058737   | AY058737 Drosophila |
| 33         | 22    | 1.4         | 4194   | 8  | MACRRL     | MACRRL              |
| 34         | 22    | 1.4         | 13653  | 2  | AC018218   | AC018218 Drosophila |
| 35         | 22    | 1.4         | 82033  | 8  | CBRG44A05  | CBRG44A05 Caenorhab |
| 36         | 22    | 1.4         | 85862  | 3  | AB011476   | AB011476 Arabidops  |
| 37         | 22    | 1.4         | 104711 | 9  | AL391988   | AL391988 Drosophila |
| 38         | 22    | 1.4         | 139581 | 2  | AP005055   | AP005055 Oryza sat  |
| 39         | 22    | 1.4         | 146388 | 2  | AP004858   | AP004858 Oryza sat  |
| 40         | 22    | 1.4         | 155152 | 2  | AC008289   | AC008289 Drosophila |
| 41         | 22    | 1.4         | 163012 | 3  | AC102289   | AC102289 Mus muscu  |
| 42         | 22    | 1.4         | 168635 | 2  | AC012388   | AC012388 Drosophila |
| 43         | 22    | 1.4         | 171972 | 3  | AC121730   | AC121730 Rattus no  |
| 44         | 22    | 1.4         | 182278 | 2  | AC121730   | AC121730 Rattus no  |
| 45         | 22    | 1.4         | 182278 | 2  | AC121730   | AC121730 Rattus no  |

#### ALIGNMENTS

RESULT 1  
LOCUS AF273766  
DEFINITION Perna canaliculus perlin precursor, mRNA, complete cds.  
ACCESSION AF273766  
VERSION AF273766.1 GI:13383377  
KEYWORDS  
SOURCE Perna canaliculus.  
ORGANISM Perna canaliculus.  
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
Mytiloidea; Mytilidae; Perna.  
REFERENCE  
1 (bases 1 to 1700)  
Scotti, P.D., Dearling, S.C., Greenwood, D.R. and Newcomb, R.D.  
Pernin: a novel, self-aggregating haemolymph protein from the New  
Zealand green-lipped mussel, Perna canaliculus (Bivalvia).

JOURNAL Mytilidae)  
Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 128 (4), 767-779  
(2001)  
MEDLINE 21186417  
PubMed 11290459  
REFERENCE 2 (bases 1 to 1700)  
AUTHORS Scotti, P. D., Dearing, S. C., Greenwood, D. R. and Newcomb, R. D.  
TITLE Direct Submission  
JOURNAL Submitted (31-MAY-2000) The Horticulture and Food Research  
Institute of New Zealand Ltd, 120 Mt. Albert Road, Auckland, New  
Zealand

| FEATURES | Location/Qualifiers           |
|----------|-------------------------------|
| source   | 1. 1700                       |
|          | /organism="Perna canaliculus" |
|          | /db_xref="taxon:38949"        |
| CDS      | 34. .1587                     |

[illegible][illegible]

|    |      |   |      |
|----|------|---|------|
| Oy | 439  | ACGAGCACAACCGAAGCGTCATTTGGACACTCAATGACTATTTTACAAAGGAGTAC        | 498  |
| Db | 532  | ACAGACACAAACCGAAGCGTCATTTGGACACTCAATGACTATTTTACAAAGGAGTAC       | 591  |
| Oy | 499  | ACCGATGCGATACCCGACGATGAATAGCGCTGTGTGTTATTTGGTCATGGAAGT          | 558  |
| Db | 592  | ACCGATGCGATACCCGACGATGAATAGCGCTGTGTGTTATTTGGTCATGGAAGT          | 651  |
| Oy | 559  | CGCCCAAGAACAGCAGCTGCTTCATCACAGCTATGAGGAATTAATCTGAGCATTT         | 618  |
| Db | 652  | CGCCCAAGAACAGCAGCTGCTTCATCACAGCTATGAGGAATTAATCTGAGCATTT         | 711  |
| Oy | 619  | GCCCATGTGACGTAAAGTCTATACACACACCAAAAGCTCTTCATCATCATGTCCAC        | 678  |
| Db | 712  | GCCCATGTGACGTAAAGTCTATACACACACCAAAAGCTCTTCATCATCATGTCCAC        | 771  |
| Oy | 679  | GGAACCATGATTTCAAAAGTTGGTTATTTGGTGACCTTGAAGTGTCTACATTTTGA        | 738  |
| Db | 772  | GGAACCATGATTTCAAAAGTTGGTTATTTGGTGACCTTGAAGTGTCTACATTTTGA        | 831  |
| Oy | 739  | GGATTTTAATGTAAAGTAGATGACACAAAGATCATCTTCACGTACGTACAGATCTACGCCAAC | 798  |
| Db | 832  | GGATTTTAATGTAAAGTAGATGACACAAAGATCATCTTCACGTACGTACAGATCTACGCCAAC | 891  |
| Oy | 799  | GGTAGCTACACAGAGATGTGATTAACCTCGGTCTAAATATGATCTCTCATGGAATTTAC     | 858  |
| Db | 892  | GGTAGCTACACAGAGATGTGATTAACCTCGGTCTAAATATGATCTCTCATGGAATTTAC     | 951  |
| Oy | 859  | CACAGATGTTGGGTGATCTGTGAGGATATTACAGATGATGACCAATGGCGTTGTCATGAA    | 918  |
| Db | 952  | CACAGATGTTGGGTGATCTGTGAGGATATTACAGATGATGACCAATGGCGTTGTCATGAA    | 1011 |
| Oy | 919  | AGGCACAGATTTCTCGTGCATCATATCTTCGCGTGAATGACAGTGTCTGGGAGCTGTAT     | 978  |
| Db | 1012 | AGGCACAGATTTCTCGTGCATCATATCTTCGCGTGAATGACAGTGTCTGGGAGCTGTAT     | 1071 |
| Oy | 979  | GGCATTACACAAAGAGACATCTTCAATAAATGGCCAAATTTGGCTGTGTGTCATAGGA      | 1038 |
| Db | 1072 | GGCATTACACAAAGAGACATCTTCAATAAATGGCCAAATTTGGCTGTGTGTCATAGGA      | 1131 |
| Oy | 1039 | CGTGCACAGAGCCATCCGAAATTTGTTCAAGAGCTAAATGTGTGTACAGCTAATACA       | 1096 |
| Db | 1132 | CGTGCACAGAGCCATCCGAAATTTGTTCAAGAGCTAAATGTGTGTACAGCTAATACA       | 1191 |
| Oy | 1099 | GAATCTACTGTTTTACATCACACATGTCGTGTGTTCTATACATTCGAAACAGACCCCTGA    | 1158 |
| Db | 1192 | GAATCTACTGTTTTACATCACACATGTCGTGTGTTCTATACATTCGAAACAGACCCCTGA    | 1251 |
| Oy | 1159 | GGATTCACACATATAGACGGCTGATCTCCAAAGATTTAAAGCTTATGTGAGGACTGTGCAT   | 1218 |
| Db | 1252 | GGATTCACACATATATAGACGGCTGATCTCCAAAGATTTAAAGCTTATGTGAGGACTGTGCAT | 1311 |
| Oy | 1219 | CATCGCATGAGTGTCAGAGCTCCATGATGGGAGATATGTCCCATGGCTCTCACCTCTTA     | 1278 |
| Db | 1312 | CATCGCATGAGTGTCAGAGCTCCATGATGGGAGATATGTCCCATGGCTCTCACCTCTTA     | 1371 |
| Oy | 1279 | GGCAGAAATTTACAGTGTCATGATGATGCGTACAGACCCCAAAACCTGGTGACTTGT       | 1338 |
| Db | 1372 | GGCAGAAATTTACAGTGTCATGATGATGCGTACAGACCCCAAAACCTGGTGACTTGT       | 1431 |
| Oy | 1339 | GATGTTATAGATGATCCCATGGCATCTGTCATCTCACTACATACCTTTGATCATCTTAT     | 1396 |
| Db | 1432 | GATGTTATAGATGATCCCATGGCATCTGTCATCTCACTACATACCTTTGATCATCTTAT     | 1491 |
| Oy | 1399 | GTTTGAAGATCTTAACGACAGTCCCTTGTGATTAATGACGGCGGACATAGAGTCCAGAGT    | 1456 |
| Db | 1492 | GTTTGAAGATCTTAACGACAGTCCCTTGTGATTAATGACGGCGGACATAGAGTCCAGAGT    | 1551 |
| Oy | 1459 | GAGAGGTTGCTGCTGTTTATAGACGGGCA 1491                              |      |
| Db | 1552 | GAGAGGTTGCTGCTGTTTATAGACGGGCA 1584                              |      |

```

RESULT 2
AC020131
LOCUS
DEFINITION AC020131 10210 bp DNA linear HTG 03-JAN-2000
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC020131
VERSION AC020131.1 GI:6664766
KEYWORDS HTG_PhasE2.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 10210)
AUTHORS Adams,M. and Venter,J.C.
JOURNAL Direct Submission
SUBMITTED (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212152 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 10210
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 3082 a 2176 c 2398 g 2554 t
ORIGIN
Query Match 1.7%; Score 26; DB 2; Length 10210;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 922 CACGATATTCCTCGATCAATATCTT 947
|||||
Db 3128 CACGATATTCCTCGATCAATATCTT 3153
|||||

RESULT 3
AC005714
LOCUS
DEFINITION AC005714 17735 bp DNA linear INV 09-NOV-2001
Drosophila melanogaster, chromosome 2R, region 58D4-58E2, BAC clone
BACR48M13, complete sequence.
ACCESSION AC005714
VERSION AC005714.10 GI:16874830
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 17735)
AUTHORS Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanaalides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Calson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostlin,D., Howland,T.J.,
Idegawa,C., Jaitai,M., Kruse,D., Li,P., Mattei,B., Mostrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Phalen,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Poulsenavong,S., Piltman,G.S., Puri,V., Richards,S., Scheler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 58D4-58E2
Unpublished
2 (bases 1 to 17735)
AUTHORS Celinker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,

```

```

TITLE
JOURNAL
COMMENT
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
location/Qualifiers
1. 17735
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="58D4-58E2"
/clone="BACR48M13 (D436)"
/clone_1lb="RC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.0)"
BASE COUNT 49614 a 39531 c 40152 g 48418 t
ORIGIN
Query Match 1.7%; Score 26; DB 3; Length 17735;
Best Local Similarity 100.0%; Pred. No. 0.0037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 922 CACGATATTCCTCGATCAATATCTT 947
|||||
Db 99256 CACGATATTCCTCGATCAATATCTT 99281
|||||

RESULT 4
AE003457
LOCUS
DEFINITION AE003457 307363 bp DNA linear INV 28-JUN-2002
Drosophila melanogaster 2R section 65 of 74 of the complete arm,
complete sequence.
ACCESSION AE003457
VERSION AE003457.2 GI:21626551
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 307363)
AUTHORS Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanaalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabot
Miklos,G.L., Abriil,J.F., Aghayani,A., An,H.J.,
Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhattacharya,P., Bolshakov,S., Borikova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Butts,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlie,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunlop,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,

```

[illegible]



```

/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBgn0034718"
complement(join(11764..14560,25247..25467))
/gene="Windpipe"
/product="CG3413-RD"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBgn0034718"
complement(join(11764..14560,21759..21807,24635..24857))
/gene="Windpipe"
/product="CG3413-RC"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBgn0034718"
complement(join(11764..14560,21759..21807,23481..23501))
/gene="Windpipe"
/product="CG3413-RA"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBgn0034718"
complement(12437..114490)
/gene="Windpipe"
/note="Windpipe gene product from transcript CG3413-RA"
/codon_start=1
/product="CG3413-PA"
/protein_id="AAM68219.1"
/db_xref="GI:21626553"
/db_xref="FLYBASE:FBan0003413"
/db_xref="FLYBASE:FBgn00034718"
Query Match 1.7%: Score 26; DB 3; Length 307363;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 922 CACGATATTCCTGCATCATATCTT 947
Db 134299 CACGATATTCCTGCATCATATCTT 134324

RESULT 5
AC108920 65256 bp DNA linear HTG 14-JUN-2002
LOCUS Mus musculus clone RP24-398E22, LOW-PASS SEQUENCE SAMPLING.
AC108920
AC108920.2 GI:21426206
HTG, HTGS_PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 65256)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-398E22
Unpublished
2 (bases 1 to 65256)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choepey,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardy,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamel,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,
Lander,E., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Menus,L., Mihov,T.,
Miyano,C.H., O'Connor,T., Nguyen,C., Nicol,R., Norbu,C.,
Peterson,K., Phukhang,P., Pierre,N., Polara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,B., Schaubov,N., Seaman,S.,
Severiy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 65256)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepey,Y., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardy,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamel,A., Karatas,A., Kells,C., Labocque,K., Lamazares,R.,
Lander,E., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Menus,L., Mihov,T.,
Miyano,C.H., O'Connor,T., Nguyen,C., Nicol,R., Norbu,C.,
Peterson,K., Phukhang,P., Pierre,N., Polara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,B., Schaubov,N., Seaman,S.,
Severiy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WMR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L21124
Center clone name: 398_E_22
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
746 845: gap of 100 bp
846 1599: contig of 754 bp in length
1600 1699: gap of 100 bp
1700 2445: contig of 746 bp in length
2446 2545: gap of 100 bp
2546 3287: contig of 742 bp in length
3288 3387: gap of 100 bp
3388 4099: contig of 712 bp in length
4100 4199: gap of 100 bp
4200 4950: contig of 751 bp in length
4951 5050: gap of 100 bp
5051 5787: contig of 737 bp in length
5788 5887: gap of 100 bp
5888 6636: contig of 749 bp in length
6637 6736: gap of 100 bp
6737 7473: contig of 737 bp in length
7474 7573: gap of 100 bp
7574 8322: contig of 749 bp in length

```

```
*      8423 8422: gap of 100 bp
*      8423 9168: contig of 746 bp in length
*      9169 9268: gap of 100 bp
*      9269 10024: contig of 756 bp in length
*      10025 10124: gap of 100 bp
*      10125 10856: contig of 732 bp in length
*      10857 10956: gap of 100 bp
*      10957 11693: contig of 737 bp in length
*      11694 11793: gap of 100 bp
*      11794 12530: contig of 737 bp in length
*      12531 12630: gap of 100 bp
*      12631 13358: contig of 728 bp in length
*      13359 13458: gap of 100 bp
*      13459 14189: contig of 731 bp in length
*      14190 14289: gap of 100 bp
*      14290 15015: contig of 726 bp in length
*      15016 15115: gap of 100 bp
*      15116 15856: contig of 741 bp in length
*      15857 15956: gap of 100 bp
*      15957 16699: contig of 743 bp in length
*      16700 16799: gap of 100 bp
*      16800 17548: contig of 749 bp in length
*      17549 17648: gap of 100 bp
*      17649 18382: contig of 734 bp in length
*      18383 18482: gap of 100 bp
*      18483 19204: contig of 722 bp in length
*      19205 19304: gap of 100 bp
*      19305 20056: contig of 752 bp in length
*      20057 20156: gap of 100 bp
*      20157 20896: contig of 740 bp in length
*      20897 20996: gap of 100 bp
*      20997 21742: contig of 746 bp in length
*      21743 21842: gap of 100 bp
*      21843 22571: contig of 729 bp in length
*      22572 22671: gap of 100 bp
*      22672 23410: contig of 739 bp in length
*      23411 23510: gap of 100 bp
*      23511 24251: contig of 741 bp in length
*      24252 24351: gap of 100 bp
*      24352 25091: contig of 740 bp in length
*      25092 25191: gap of 100 bp
*      25192 25940: contig of 749 bp in length
*      25941 26040: gap of 100 bp
*      26041 26788: contig of 748 bp in length
*      26789 26888: gap of 100 bp
*      26889 27619: contig of 751 bp in length
*      27620 27719: gap of 100 bp
*      27720 28356: contig of 617 bp in length
*      28357 28456: gap of 100 bp
*      28457 29185: contig of 749 bp in length
*      29186 29285: gap of 100 bp
*      29286 30025: contig of 740 bp in length
*      30026 30125: gap of 100 bp
*      30126 30880: contig of 755 bp in length
*      30881 30980: gap of 100 bp
*      30981 31713: contig of 733 bp in length
*      31714 31813: gap of 100 bp
*      31814 32567: contig of 754 bp in length
*      32568 32667: gap of 100 bp
*      32668 33400: contig of 733 bp in length
*      33401 33500: gap of 100 bp
*      33501 34251: contig of 751 bp in length
*      34252 34351: gap of 100 bp
*      34352 35103: contig of 752 bp in length
*      35104 35203: gap of 100 bp
*      35204 35945: contig of 742 bp in length
*      35946 36045: gap of 100 bp
*      36046 36777: contig of 732 bp in length
*      36778 36877: gap of 100 bp
*      36878 37626: contig of 749 bp in length
*      37627 37726: gap of 100 bp
*      37727 38471: contig of 745 bp in length
*      38472 38571: gap of 100 bp
```

```
*      38572 39315: contig of 744 bp in length
*      39316 39415: gap of 100 bp
*      39416 40152: contig of 737 bp in length
*      40153 40252: gap of 100 bp
*      40253 40988: contig of 736 bp in length
*      40989 41088: gap of 100 bp
*      41089 41840: contig of 752 bp in length
*      41841 41940: gap of 100 bp
*      41941 42702: contig of 762 bp in length
*      42703 42802: gap of 100 bp
*      42803 43561: contig of 759 bp in length
*      43562 43661: gap of 100 bp
*      43662 44398: contig of 738 bp in length
*      44400 44499: gap of 100 bp
*      44500 44520: contig of 721 bp in length
*      45221 45320: gap of 100 bp
*      45321 46062: contig of 742 bp in length
*      46063 46162: gap of 100 bp
*      46163 46904: contig of 742 bp in length

Query Match
Best Local Similarity 100.0%; Score 25; DB 2; Length 65256;
Matches 25; Conservative 0; Mismatches 0; Gaps 0;

*      68 ATGACGACCATGATGATGATGATGA 92
*      Db 8829 ATGACGACCATGATGATGATGATGA 8853
*      |||||||||||||||||||||||||
*      AC019858 67759 bp DNA linear HTG 03-JAN-2000
*      LOCUS AC019858/c
*      DEFINITION
*      AC019858
*      AC019858
*      AC019858.1 GI:665039
*      VERSION
*      HTG: PHASE2.
*      KEYWORDS
*      Drosophila melanogaster.
*      SOURCE
*      ORGANISM
*      Drosophila melanogaster.
*      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
*      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
*      Ephydroidea; Drosophilidae; Drosophila.
*      REFERENCE
*      1 (bases 1 to 67759)
*      AUTHORS
*      Adams,M. and Venter,J.C.
*      TITLE
*      Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
*      Rockville, MD, USA
*      JOURNAL
*      COMMENT
*      This sequence was identified as CDM:10211184 by the submitter.
*      For more information on this record e-mail to fly@celera.com.
*      * NOTE: This is a 'working draft' sequence.
*      * This sequence will be replaced
*      * by the finished sequence as soon as it is available and
*      * the accession number will be preserved.
*      Location/Qualifiers
*      FEATURES
*      1..67759
*      source
*      /organism="Drosophila melanogaster"
*      /db_xref="taxon:7227"
*      BASE COUNT 18034 a 15223 c 15520 g 18982 t
*      ORIGIN
*      Query Match
*      Best Local Similarity 100.0%; Score 25; DB 2; Length 67759;
*      Matches 25; Conservative 0; Mismatches 0; Indels 0;
*      Oy 68 ATGACGACCATGATGATGATGATGA 92
*      Db 33276 ATGACGACCATGATGATGATGATGA 33252
*      |||||||||||||||||||||||||
*      RESULT 7
*      AC010057 192588 bp DNA linear INV 13-JUN-2002
*      LOCUS AC010057
*      DEFINITION Drosophila melanogaster 3L BAC RP98-26C18 (Roswell Park Cancer
```



Identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot>

action.html

# FEATURES

Source Location/Qualifiers

1..192588  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="3L"

BASE COUNT 52876 a 43443 c 42611 g 53658 t

## ORIGIN

Query Match 1.7%; Score 25; DB 3; Length 192588;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 ATGACGACATGATGATGATGATGA 92  
|||||  
Db 171304 ATGACGACATGATGATGATGATGA 171328

## RESULT 8

AC116539

LOCUS

DEFINITION

Drosophila melanogaster clone RP98-18EL, WORKING DRAFT SEQUENCE, 3

ACCESSION

AC116539 2 GI:21930209

HTG: HTGS-PHASE1 HTGS-DRAFT; HTGS-FULLTOP; HTGS-ACTIVEPIN.

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 207018)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T.,

Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D.,

Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,

Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,

Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

DeLaney,K.R., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Delaney,K.R., Delgado,O., Degen,A.L., Ding,Y., Dinn,H.H.,

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,

Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,

Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,

Gorelli,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A., Hernandez,J.,

Hernandez,O., Hodgson,A., Hughes,M., Hollaway,C., Hollins,B.,

Homs,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolyvel,S., Joudan,S.,

Karlsson,B., Kelly,S., Khan,O., King,L., Kovach,C.,

Krautovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Louise,H.,

Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Lunde,R., Ma,J.,

Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

Massery,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metker,M.,

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okunou,G.,  
Oreguena,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherrer,S., Scott,G., Shen,H., Shoshitani,N., Sisson,I.,  
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,  
Williams,G., Williamson,A., Wleczek,R., Woodson,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G., and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 207018)

Worley,K.C.

Direct Submission

Submitted (29-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 207018)

Worley,K.C.

Direct Submission

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jul 23, 2002 this sequence version replaced gi:19807671.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: DRML

Center clone name: RP98-18EL

----- Summary Statistics

Chemistry: Dye-Primer Method; Infinity% of reads

Assembly: Dye-terminator Big Dye; Infinity% of reads

Consensus quality: 21917 bases at least Q40

Consensus quality: 224732 bases at least Q30

Consensus quality: 23117 bases at least Q20

Estimated insert size: 207648; sum-of-contigs estimation

Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 6420: contig of 6420 bp in length

\* 6421 6530: gap of unknown length

\* 91457 91456: contig of 84936 bp in length

\* 91457 91556: gap of unknown length

\* 91557 207018: contig of 115466 bp in length.

Location/Qualifiers

1..207018

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="RP98-18EL"

BASE COUNT 5739 a 45880 c 45706 g 57469 t 224 others

## ORIGIN

Query Match 1.7%; Score 25; DB 2; Length 207018;  
Best Local Similarity 100.0%; Pred. No. 0.015;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



## MRNA

/db\_xref="FLYBASE:FBan0012189"  
join(26843..26959,27018..28375,28430..28712,28775..29198,  
29261..29314,29380..30067,30132..30455)  
/gene="CG12189"  
/product="Cr9555"  
/db\_xref="FLYBASE:FBan0012189"  
join(26924..26959,27018..28375,28430..28712,28775..29198,  
29261..29314,29380..30067,30132..30216)  
/gene="CG12189"  
/note="CG12189 gene product"

## CDS

/codon\_start=1  
/protein\_id="AA047401.1"  
/db\_xref="GI:7291986"  
/db\_xref="FLYBASE:FBan0012189"  
/db\_xref="FLYBASE:FBan0035150"  
/translation="MTDEDNGFSEMGGEFEAKSKLEQFAAASDPFRKSDLPFGIS  
IFVNGRTDPSADELKRIMVHGTFHHYERSHTTYIIASVLPVKRNMNLKFTSAK  
WVYDCLERKRIYDYPYLLYNQTSQPMIFGPKDGNESKSDVEPPKDAVEY  
DSTKDETMELGILKLNQAVATSPEKASASKITNSTSNSTASTARADPNF  
LSEFYKNSRLHIITATGAGKQVYRLROKHTGQFPRETLKSLNSHNLERYM  
HIDMDFEVSGLRTPRLRGLPIAVTHSGKGNATVPVHPQADRKALDELFAORF  
HHFHDGKAERKSGDFKMSLSEASCSYEARERKGRNMGALKICPELKTIPY  
DEFGYEVAFITDYAOTYNTIEAVSGDEMFVELDIAHELVDMVMAFVSLREYV  
SKTGCSAGVAGNKLARMAKREKPNQGLDSSMDLAELVMAFVSLREYV  
SRTKQAGLNGGDVONTTLEKMEVGLKLGONLQNGCIGDDRLAELAEJOIRKVS  
EMNGIRFPTNSVECEPFCOLSEVTRVIEIRRKARISIKLIVRAAFAVETSRY  
GHGVDIIRKSLIKYATDQVNTITVVLDMQADVPDELGRGKIHTRLEDNEY  
LRKNNIKEMEGKSEMRKDKPIQGAVDGSGIDGDKNPKLVEENKPREPRANYSM  
LITAAVSRKSVTEEDRSRGTSKPIITRPLSITPRLDEVLQAQLEPDLRLVLRNREH  
CIAEYRSQDITVADIDDLDPFVNPHTIKLISHPEMPELIMQNDYDLNDVSEEV  
PKNDVDLILKOVSRMINKDLDHVCVMKWKCRILIMKRSSCCNHVAKIIESIO  
NOMLTREGSLTFEYIKICS"

## gene

/complement(44992..52784)  
/gene="CG17129"  
/map=61C8-61C9  
/db\_xref="FLYBASE:FBan0017129"  
/db\_xref="FLYBASE:FBan0035151"  
complement(join(4992..5007,3,50495..51310,51776..52309,  
52379..52784))  
/gene="CG17129"  
/product="Cr38022"  
/db\_xref="FLYBASE:FBan0017129"  
/db\_xref="FLYBASE:FBan0035151"  
complement(join(4992..5007,3,50495..51310,51776..52227))  
/gene="CG17129"  
/note="CG17129 gene product"

## CDS

/codon\_start=1  
/protein\_id="AA047402.1"  
/db\_xref="GI:7291987"  
/db\_xref="FLYBASE:FBan0017129"  
/db\_xref="FLYBASE:FBan0035151"  
/translation="MAKWLRLDEVGGEKSOODKMESESPAOENOKOKVLANPSV  
VDCSEVEKSVFRDMETLDVESVLPNSLPGSGOTDGETTDLAKDVLNKLDE  
ASLRKRNREEDSPSHNAKORSIKELFLPKKSGOTDGETTDLAKDVLNKLDE  
IRVRLDLEOVDSPEPKALGTAKELRLPKKOCNSLOPKTMDXONPISVTSST  
NSPNFPAFAPIITTYOINRLPSPAPATSSNPATNSPVSOTSGILPAVTF  
HDENPRLPSVTTVTYNSRYVDPDRPANSWYTPRNHNPATVPVPTSTSTYOPT  
SSSYGONPVAIVLSSGYIRGPPDPSANNAANPAIGSRPDLNTIRMAIDVSPFAC  
ONNREAYHVLNQLRHSLOTGANSRFOAPSEMPAIKDKREKREKIQDDANAFNWF"

Query Match 1.7%: Score 25; DB 3; Length 302527;  
Best Local 25; Similarity 100.0%; Pred. No. 0.016;

Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 ATGAGACCATGATGATGATGATGA 92  
|||||

DB 197335 ATGAGACCATGATGATGATGATGA 197359

RESULT 10

AF231797 476 bp DNA linear PLN 24-JUL-2001  
LOCUS AF231797  
DEFINITION Pinus taeda clone LC15112 microsatellite sequence.  
ACCESSION AF231797  
VERSION AF231797.1 GI:15004298  
KEYWORDS  
SOURCE Pinus taeda  
ORGANISM Pinus taeda  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus;

REFERENCE 1 (bases 1 to 476)  
AUTHORS Elsik,C.G. and Williams,C.G.  
TITLE Microsatellite organization in a large plant genome  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 476)  
AUTHORS Elsik,C.G. and Williams,C.G.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2000) Department of Forest Science, Texas A&M  
University, Room 305 HFSB, College Station, TX 77843-2135, USA  
Location/Qualifiers

FEATURES  
source  
1..476  
/organism="Pinus taeda"  
/db\_xref="taxon:3352"  
/clone="LC15112"  
repeat\_region  
1..476  
/note="microsatellite; trinucleotide repeat"

BASE COUNT 187 a 50 c 93 g 144 t 2 others  
ORIGIN  
1..476  
/rpt\_type=tandem

Query Match 1.6%: Score 24; DB 8; Length 476;  
Best Local 24; Similarity 100.0%; Pred. No. 0.033;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TGACGACCATGATGATGATGATGA 92  
|||||

DB 287 TGACGACCATGATGATGATGATGA 310

RESULT 11  
LOCUS AF172408 508 bp DNA linear PLN 06-AUG-2000  
DEFINITION Pinus taeda clone LC15112 microsatellite sequence.  
ACCESSION AF172408  
VERSION AF172408.1 GI:9719232  
KEYWORDS  
SOURCE Pinus taeda  
ORGANISM Pinus taeda  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus;

REFERENCE 1 (bases 1 to 508)  
AUTHORS Elsik,C.G. and Williams,C.G.  
TITLE Microsatellite recovery from the low-copy component of a large  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 508)  
AUTHORS Elsik,C.G. and Williams,C.G.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUL-1999) Forest Science, Texas A&M University, HFSB  
Room 305, College Station, TX 77843-2135, USA  
Location/Qualifiers

FEATURES  
source  
1..508  
/organism="Pinus taeda"  
/db\_xref="taxon:3352"  
/clone="LC15112"  
repeat\_region  
277..474  
/note="microsatellite"

BASE COUNT 195 a 61 c 100 g 149 t 3 others  
ORIGIN  
1..6%: Score 24; DB 8; Length 508;

Best Local Similarity 100.0%; Pred. No. 0.034; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TGACGACCATGATGATGATGA 92  
|||||  
Db 286 TGACGACCATGATGATGATGA 309

RESULT 12  
AL590633/c 179289 bp DNA linear ROD 05-APR-2002  
LOCUS Mouse DNA sequence from clone RP23-211023 on chromosome X, complete  
DEFINITION sequence.  
ACCESSION AL590633  
VERSION AL590633.8 GI:20068454  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1  
Wall, M.  
Direct Submission  
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Apr 7, 2002 this sequence version replaced gi:15617304.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submissions corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep  
from the RPI-23 mouse PAC library  
constructed by the group of Pletier de Jong.  
for further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pPAC3.0.

FEATURES  
source  
1..179289  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-211023"  
/clone\_11b="RPI-23"  
BASE COUNT 56940 a 36892 c 36607 g 48850 t  
ORIGIN

Query Match 1.6%; Score 24; DB 10; Length 179289;  
Best Local Similarity 100.0%; Pred. No. 0.062;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 CACCATGACCATGATGATGAT 87  
|||||  
Db 175276 CACCATGACCATGATGATGAT 175253

RESULT 13  
AL845279 209173 bp DNA linear HTG 09-AUG-2002  
LOCUS Mus musculus chromosome X clone RP23-2E13, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 12 unordered pieces.

ACCESSION AL845279  
VERSION AL845279.1 GI:22204974  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 209173)  
McLay, K.  
Direct Submission  
Submitted (02-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Aug 2, 2002 this sequence version replaced gi:22204974.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submissions corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep  
from the RPI-23 mouse PAC library  
constructed by the group of Pletier de Jong.  
for further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pPAC3.0.

REFERENCE 1  
Wall, M.  
Direct Submission  
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
On Apr 7, 2002 this sequence version replaced gi:15617304.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submissions corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at  
http://www.sanger.ac.uk/projects/C\_elegans/wormpep  
from the RPI-23 mouse PAC library  
constructed by the group of Pletier de Jong.  
for further details see http://www.chori.org/bacpac/home.htm  
VECTOR: pPAC3.0.

FEATURES  
source  
1..209173  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-2E13"  
/clone\_11b="RPI-23"  
1..11646  
/note="assembly\_fragment:02548.0"  
11747..27085  
/note="assembly\_fragment:02548.1"

misc\_feature  
misc\_feature

Assembly program: XGAP4, version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 207186 bases at least Q40  
Consensus quality: 207812 bases at least Q30  
Consensus quality: 207812 bases at least Q20  
Insert size: 208073; sum-of-contigs  
Quality coverage: 9.32x in Q20 bases; sum-of-contigs Quality coverage: 9.44x in Q20 bases; agarose-1p

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 11646: contig of 11646 bp in length  
\* 11647 11746: gap of 100 bp  
\* 11747 27085: contig of 15339 bp in length  
\* 27086 27185: gap of 100 bp  
\* 27186 30891: contig of 3706 bp in length  
\* 30892 30991: gap of 100 bp  
\* 30992 62508: contig of 31517 bp in length  
\* 62509 62608: gap of 100 bp  
\* 62609 68419: contig of 5811 bp in length  
\* 68420 68519: gap of 100 bp  
\* 68520 126394: contig of 57875 bp in length  
\* 126395 126494: gap of 100 bp  
\* 126495 142463: contig of 15969 bp in length  
\* 142464 142563: gap of 100 bp  
\* 142564 145884: contig of 3321 bp in length  
\* 145885 145984: gap of 100 bp  
\* 145985 194833: contig of 48849 bp in length  
\* 194834 194933: gap of 100 bp  
\* 194934 198509: contig of 3576 bp in length  
\* 198510 198609: gap of 100 bp  
\* 198610 206634: contig of 8025 bp in length  
\* 206635 206734: gap of 100 bp  
\* 206735 209173: contig of 2439 bp in length.

Location/Qualifiers  
1..209173  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-2E13"  
/clone\_11b="RPI-23"  
1..11646  
/note="assembly\_fragment:02548.0"  
11747..27085  
/note="assembly\_fragment:02548.1"

```
misc_feature 27186..30891
/note="assembly-fragment:02412
fragment_chain:1"
misc_feature 30992..62508
/note="assembly-fragment:02521
fragment_chain:1"
misc_feature 62609..68419
/note="assembly-fragment:02028
fragment_chain:1"
misc_feature 68520..126394
/note="assembly-fragment:00458
fragment_chain:1"
misc_feature 126495..142463
/note="assembly-fragment:00628
fragment_chain:1"
misc_feature 142564..143884
/note="assembly-fragment:02995
fragment_chain:1"
misc_feature 145985..194833
/note="assembly-fragment:03312
fragment_chain:1"
misc_feature 194934..198509
/note="assembly-fragment:01240
fragment_chain:1"
misc_feature 198610..206634
/note="assembly-fragment:00991
fragment_chain:1"
misc_feature 206735..209173
/note="assembly-fragment:01353
fragment_chain:1"
misc_feature 206735..209173
/note="assembly-fragment:01353
fragment_chain:1"
vector_side:right"
BASE COUNT 62758 a 43215 c 42828 g 59259 t 1113 others
ORIGIN
Query Match 1.6%; Score 24; DB 2; Length 209173;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 64 CACCATGACGACATGATGATGAT 87
|||||
Db 192630 CACCATGACGACATGATGAT 192653

RESULT 14
BRUCSP/LOCUS 1817 bp DNA linear BCT 26-APR-1993
DEFINITION B.abortus BCSp31 gene encoding a 31-KDa cell surface protein,
complete cds.
ACCESSION M20404
VERSION M20404.1 GI:144104
KEYWORDS cell surface protein.
SOURCE B.abortus (strain 19) DNA, clone PBA31-R7.
ORGANISM Brucella mellitensis biovar Abortus
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; Brucella.
REFERENCE 1 (bases 1 to 1817)
Mayfield,J.E., Bricker,B.J., Godfrey,H., Crosby,R.M., Knight,D.J.,
Halling,S.M., Balinsky,D., and Tabatabai,L.B.
The cloning, expression, and nucleotide sequence of a gene coding
for an immunogenic Brucella abortus protein
JOURNAL Gene 63 (1), 1-9 (1988)
MEDLINE 88255848
PUBMED 3133283
FEATURES
Source location/Qualifiers
1..1817
/organism="Brucella mellitensis biovar Abortus"
/db_xref="taxon:235"
/note="Biovar Abortus"
585..1354
/note="cell surface protein precursor"
/codon_start=1
/transl_table=11

/protein_id="AA22993.1"
/db_xref="GI:144105"
/translation="MKFGSKIRLVAAVAGAIAGSFAVAQAPTEFRIGTGACT
YPIIGLIVANASIGAGEKGPGLVATAVSNGSVANIKASGALSGFQSDVAYWA
YNGTGLVDGKVEDIRLRLATLYPEIRIHVAROKANIKSVADLKGKRSIDEPGSGT
VDARIVAEVGLTEDIDIKAEHLKPPAGBRLLKDGALDYPVPGYPTGASISLAI
ISLVISGPEADKILKYSFSPKDVYPAGAVPEPTTVAQAVTSAKOPDILY
NITKLVNEDTRKALDAGHAKKLIKLSATSSGIPLHPGAEFRYEAGVLK"
sig_peptide
mat_peptide
BASE COUNT 382 a 468 c 539 g 428 t
ORIGIN 515 bp upstream of HindIII site.

Query Match 1.5%; Score 22; DB 1; Length 1817;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1331 ACCTGTGATGTTATAGATGA 1352
|||||
Db 1406 ACCTGTGATGTTATAGATGA 1385

RESULT 15
AE009521/LOCUS 9960 bp DNA linear BCT 28-DEC-2001
DEFINITION Brucella mellitensis strain 16M chromosome I, section 78 of 195 of
the complete sequence.
ACCESSION AE009521
VERSION AE009521.1
KEYWORDS
SOURCE AE009521.1 GI:17982737
ORGANISM Brucella mellitensis.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiales group;
Brucellaceae; Brucella.
REFERENCE 1 (bases 1 to 9960)
DelVecchio,V.G., Redkar,R.J., Patra,G., and Mujer,C.
Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
JOURNAL 11756688
PUBMED 11756688
REFERENCE 2 (bases 1 to 9960)
DelVecchio,V.G., Redkar,R.J., Patra,G., and Mujer,C.
Direct Submission
JOURNAL Submitted (13-NOV-2001) Institute of Molecular Biology and
Medicine, University of Scranton, Scranton, PA 18510, USA
3 (bases 1 to 9960)
Elzer,P.H. and Hagius,S.
Direct Submission
JOURNAL Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA
4 (bases 1 to 9960)
Kaprali,V., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A.,
Lykidis,A., Rezik,G., Jablonski,L., Larsen,N., D'Souza,M.,
Bernal,A., Mazur,M., Goltsman,E., Selkov,E., Haselkorn,R.,
Kyrpides,N., and Overbeek,R.
Direct Submission
JOURNAL Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
Park Drive, IL 60612, USA
5 (bases 1 to 9960)
Letesson,J.-U.
Direct Submission
JOURNAL Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire,
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
61 rue de Bruxelles, Namur 5000, Belgium
6 (bases 1 to 9960)
O'Callaghan,D.
```



TITLE Direct Submission  
JOURNAL Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue  
Kennedy, Nimes 30900, France  
FEATURES Location/Qualifiers  
source  
1. 9960  
/organism="Brucella melitensis"  
/strain="16M"  
/db\_xref="taxon:29459"  
/chromosome="I"  
104. .1093  
/gene="BMEI0796"  
104. .1093  
/gene="BMEI0796"  
104. .1093  
/codon\_start=1  
/transl\_table=11  
/product="31 kDa IMMUNOGENIC PROTEIN PRECURSOR"  
/protein\_id="AAL51977.1"  
/db\_xref="GI:17982738"  
/transl\_table=11  
/transl\_table="1"KFGSKRLRLAANAAGATAGASFAVAQAFTPRFGTGTACT  
VTPGGGLANALSGAGKGVGLVATIAVSSNGSVANTINAKSGALESGFTQSDVAVYA  
VNGTGLYDGRKVEDLRLATLYPETIHHVARKDANKSVADLKGRVSLDEPGSGT  
VNAVIVLSEAYGLTEDDIKAEHLKPGAGEKLGALDAYFEVGGYPTGATISLAI  
ISLVISGPEADKILEKYSFESKDVVAGAKDVATPTLAAVAAQWTSAPDPLIY  
NITVLMNEDRKALDAGHAKGLIKLDSYSLGIPHPAGEREYKEAGLK"  
1330. .3378  
/gene="BMEI0797"  
1330. .3378  
/gene="BMEI0797"  
/codon\_start=1  
/transl\_table=11  
/product="TRANSPORTER"  
/protein\_id="AAL51978.1"  
/db\_xref="GI:17982739"  
/transl\_table="1"MTBSONAKLAPMELDEVKARELEEKDESEIHFRLAPVARIY  
TLILSLFHYTAGFGLPEPMIRHIGLHAFVGLVETFPFSRKGVDEPKSLP  
LGISVDMGLATIAVAVIHPILIPDLDAFRGNTSTDVLSLLIILLETFS  
VGMPLITISVLEMLYALXGSMGIIHPATASOLVHLITGTCIGTIAAGV  
VHFVLRGPVATRICGILFDCAAWAGRPAGAPAKVYSRGSALRGYISGVANV  
TYSGLTTPAMTRICGYPAPAAVSASSGCGTPPIMAAETIETLNIPYTTIL  
AIVAPAFMHFQVLMQVHFPAKKTGRLGRTKEEMPKALKRWPTIIPVLAIVL  
LSCGTPYLAATNGTILCLAVGLAPFRKMTIGEVFGLDGGAKTALAGAAATGVI  
VGVVLTGVRKISTIVSTRADLATFEGTISVSWFAQTITLFTLVMAAKATGII  
GCGITPTANNIIMATIAAPALGGLVETVAHFVYVGLADITPVALAAVAAAM  
AGADPFKGTGAFRLGKVLVPPVVFVPSLLLVTSNMPDFIAGFCVIGITV  
GAALSGFPLVTRKIMENVLLIFAMLLVABEITYSSIVGLIILLPVVRLHVASRRP  
3541. .3927  
/gene="BMEI0798"  
3541. .3927  
/gene="BMEI0798"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical Protein"  
/protein\_id="AAL51979.1"  
/db\_xref="GI:17982740"  
/transl\_table="1"MLSEVOAMEIVQIRISSVGGFKLYMVEFVTEGGERITVLEND  
TKELRDEVIARRAATKIDGMAGACEGIEPDSLITLTPRSARRADAELEQLDDG  
LEDTPASDPVSVTSASIPASADPKS"  
complement(4034. .6226)  
/gene="BMEI0799"  
complement(4034. .6226)  
/gene="BMEI0799"  
/BC\_number="5.4.99.2"  
/codon\_start=1  
/transl\_table=11  
/product="METHYLALONYL-COA MUTASE"  
/protein\_id="AAL51980.1"  
/db\_xref="GI:17982741"  
/transl\_table="1"WRCLAPDQASIMDDGDEPTKTRADWETLAELKLRPADSLV  
HTEGIDVYKPLITDDLEGIGLTLGFEPLRGRATVYAGRPWTIRQYAGSTAE  
ESNAFTKRALAGGQGVVAADLATHRGDSDHPREVGGVKAIVADISVEDMKILD  
GIPLEKISVSMNGNAVIPLANIVAGEQGVPRALGCTIIONDILKEFWRVNTYIY  
PPEPSMRLIADIIVTAOEMPKFNSISISGTHOEGATLVQELAFTLADREVYRAA

LKGLVNDPFAGRLSFFAIGNFMEITAKIARAALMSIMKEFDPKRGSLMURTH  
CQFSVSLQEDPDYNNIVTFAEAMSAALGQTSOLHTNSDEDAIALPTEFSARIANT  
QILIONETGATKAVDPLAGSYIESTINMELAEKMALEIEVETLGAABSGLPKR  
LIEEATRRQVALEKGEVEIVGVNKFRLDEDEIDILEIDNSVARGOJARLNRKEM  
RNRKQVAALEAEKARSGEENLLAAVEASRARATVGEISDPMARGLDHAAPKY  
VKRYGTAAKDEEYQTLVDRLGEFAKQAGKPRVYAKLGGQDRGAKIIASFGD  
IGFEVLAPLFPOTPEADMAIKARVHVLGVSILAAQKTLPLQVETLRKKGADNII  
VVCQGVPRDYOYLDFDHGVAVFGPTNIVLDCARAVLDMECKLRNL"  
complement(6233. .8236)  
/gene="BMEI0800"  
complement(6233. .8236)  
/gene="BMEI0800"  
/BC\_number="6.4.1.3"  
/codon\_start=1  
/transl\_table=11  
/product="PROPIONYL-COA CARBOXYLASE ALPHA CHAIN,  
MITOCHONDRIAL"  
/protein\_id="AAL51981.1"  
/db\_xref="GI:17982742"  
/transl\_table="1"MIKTLILNNGELACRVIRKAKKGIATVAVYSADARNALHYK  
ADEVHIGPAPSNOSTYIVDKILAIKEIGADAVHPGGLSENPRFAALAAVNTF  
IGPPVNAITAMGDKITSKRLAAEAGSVTPGMGLIEDADEAVRIAGSITGPMYKAS  
AGGCGKMRILAMDEARREGFQLSNEAKSSFQDDRIEIKETVOPRHILEIOLGDOH  
GNVYVYLGRECSIORRNQVIEAPSPLEADETKAMGBOVALAAVGYAGTVEF  
IVDGNRFYFLEMMNTRLOVEHPVETLITGIDLVEMIRVAGSKELFAQADVNLGMA  
IESRLAEDPYRNFLPSIGRLTRYRPVPEGNNPGTVIRNDTGVEFGGKISMYDPMI  
AKLTGWPDRISALIDAMGHALDAFEVEIGHNLFELVANDHDFREGALTTAFLAE  
YPDGSGVKCEDDARTLLAAVAEINLVAORROTOIGRSLSPOKHSIANDVVTLDGY  
SLPVRIAEGEGGTTINFIDGSLPIASDMHQSOLGSGFTGKPIAVKYSRSGTMR  
RMGRMDVAHVRRPVRAELAKLMPVRLPPTSKMLCPMGVITSLVVDGFTVAGQ  
PLATEAMKENVLRAERRATVKTITAGSSVLAVDELIMEF"  
complement(8321. .9853)  
/gene="BMEI0801"  
complement(8321. .9853)  
/gene="BMEI0801"  
/BC\_number="6.4.1.3"  
/codon\_start=1  
/transl\_table=11  
/product="PROPIONYL-COA CARBOXYLASE BETA CHAIN"  
/protein\_id="AAL51982.1"  
/db\_xref="GI:17982743"  
/transl\_table="1"MOELBQLEARRAARLGGCORIDAGHCKGLTARERIEVLD  
EGSPEEDMIVYHRCDFEGMAQKVAGGVVTGNGTINGRQVVFSDPTVLGSLSE  
THAOKICKIMDAVAKNGAPVIGLNDSSYMFVTPDVVKVINEIVTAELELGAATHT  
SVINGPAGAGVAYSPAMTDFIPVNDSSYMFVTPDVVKVINEIVTAELELGAATHT  
KKSVDAGAYENDIEALEQVRIILDFELNLNRPVPIRFFDPPARLEMLDTLAVDS  
ATKPYDMKELILADDEGDFEILQEAFLNLTGFIIMDGOYGVAAVANOPLYACLD  
IDASRKARVFRCDAFNIPILTLVDVGPFGPAGQAGVIGAKLFLAVSQATVP  
MVTILTRKAVGAGVDMASKHIGADINAMPTAIAVMAGKTEILYRSELDGEKI  
AARTKEYEERFANPEVAERGFIDEVIMPHSSRRRIARAPASLRNQGAPMKKHDTI  
PL"

Search completed: November 29, 2002, 12:42:25  
Job time : 5648 secs

BASE COUNT 1917 a 2822 c 2934 g 2287 t  
ORIGIN  
Query Match 1.5%: Score 22; DB 1; Length 9960;  
Best Local Similarity 100.0%: Pred. No. 0.77;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1331 ACCTGGGATGATTAATAGCA 1352  
|||||  
DB 945 ACCTTGGTATGTTATAGATCA 924

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 08:14:04 ; Search time 270 Seconds

(without alignments)  
12436.035 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491

Sequence: 1 gaggggagcagtgtaacga.....gctgtgtatagagcggca 1491

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq.101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1          | 1491  | 100.0       | 1491   | 21 | AAAA7150    |
| 2          | 1491  | 100.0       | 1491   | 21 | AAAA7151    |
| 3          | 26    | 1.7         | 3211   | 23 | ABL06474    |
| 4          | 26    | 1.7         | 6398   | 23 | ABL10542    |
| 5          | 25    | 1.7         | 2171   | 23 | ABL23878    |
| 6          | 25    | 1.4         | 1599   | 21 | AAAC39999   |
| 7          | 21    | 1.4         | 3309   | 23 | ABL15963    |
| 8          | 21    | 1.4         | 4909   | 16 | AAAT18924   |
| 9          | 21    | 1.4         | 6699   | 23 | ABLI15962   |

|    |    |     |       |    |           |                       |
|----|----|-----|-------|----|-----------|-----------------------|
| 10 | 21 | 1.4 | 40023 | 24 | ABLI1954  | Human solute carri    |
| 11 | 20 | 1.3 | 65    | 22 | AAH45894  | C parvum coding se    |
| 12 | 20 | 1.3 | 461   | 24 | ABLI93659 | Arabidopsis thalia    |
| 13 | 20 | 1.3 | 706   | 21 | AAAC49072 | Arabidopsis thalia    |
| 14 | 20 | 1.3 | 708   | 21 | AAAC34259 | Arabidopsis thalia    |
| 15 | 20 | 1.3 | 879   | 23 | AAAG7497  | DNA encoding novel    |
| 16 | 20 | 1.3 | 1774  | 21 | AAAC42052 | Arabidopsis thalia    |
| 17 | 20 | 1.3 | 2177  | 23 | ABLI25286 | Arabidopsis thalia    |
| 18 | 19 | 1.3 | 34    | 21 | AAH60434  | Plasmodium falciparum |
| 19 | 19 | 1.3 | 64    | 22 | AAH45895  | C parvum coding se    |
| 20 | 19 | 1.3 | 87    | 19 | AAV47595  | Rhodococcus halobal   |
| 21 | 19 | 1.3 | 104   | 17 | AAAT33459 | Ich-2 gene 5' prim    |
| 22 | 19 | 1.3 | 104   | 22 | ABA72619  | Human foetal liver    |
| 23 | 19 | 1.3 | 104   | 22 | ABA73973  | Human foetal liver    |
| 24 | 19 | 1.3 | 104   | 22 | ABA38329  | Probe #16795 for g    |
| 25 | 19 | 1.3 | 104   | 22 | ABA39059  | Probe #17525 for g    |
| 26 | 19 | 1.3 | 104   | 22 | AAK21048  | Human brain expres    |
| 27 | 19 | 1.3 | 104   | 22 | AAK22426  | Human brain expres    |
| 28 | 19 | 1.3 | 104   | 22 | AAK47200  | Human bone marrow     |
| 29 | 19 | 1.3 | 104   | 22 | AAK48593  | Human bone marrow     |
| 30 | 19 | 1.3 | 104   | 22 | AAI26272  | Probe #16205 for g    |
| 31 | 19 | 1.3 | 104   | 22 | AAI53036  | Probe #21722 used     |
| 32 | 19 | 1.3 | 104   | 22 | AAI54422  | Human genome-deriv    |
| 33 | 19 | 1.3 | 104   | 24 | ABSI21403 | Human genome-deriv    |
| 34 | 19 | 1.3 | 104   | 22 | ABSI22354 | Human foetal liver    |
| 35 | 19 | 1.3 | 142   | 22 | ABSI3028  | Human brain expres    |
| 36 | 19 | 1.3 | 142   | 22 | AAK21460  | Human bone marrow     |
| 37 | 19 | 1.3 | 142   | 22 | AAK47621  | Human bone marrow     |
| 38 | 19 | 1.3 | 179   | 20 | AAI53454  | Probe #22140 used     |
| 39 | 19 | 1.3 | 179   | 20 | AAI53454  | Otopouche NP CDNA     |
| 40 | 19 | 1.3 | 213   | 21 | AAI99097  | MULTICLONING SITE     |
| 41 | 19 | 1.3 | 213   | 21 | AAI99097  | DNA fragment of pg    |
| 42 | 19 | 1.3 | 240   | 21 | AAI61541  | Nucleotide sequenc    |
| 43 | 19 | 1.3 | 432   | 21 | AAI61541  | Arabidopsis thalia    |
| 44 | 19 | 1.3 | 439   | 24 | ABLI94190 | Arabidopsis thalia    |
| 45 | 19 | 1.3 | 465   | 22 | AAI10206  | Human breast cance    |

## ALIGNMENTS

RESULT 1  
AAAA7150  
ID AAAA7150 standard; DNA; 1491 BP.  
XX  
AC AAAA7150:  
XX  
DT 03-OCT-2000 (first entry)  
XX  
DE DNA encoding a serine protease inhibitor protein.  
XX  
KW Serine protease inhibitor; green-flipped mussel; anti-thrombin;  
KW divalent metal cation binding activity; dietary supplement;  
KW anticoagulant; ss.  
XX  
OS Perna canaliculus.  
XX  
PN WO200039165-A1.  
XX  
PD 06-JUL-2000.  
XX  
PE 23-DEC-1999; 99WO-NZ00227.  
XX  
PR 23-DEC-1998; 98NZ-0333568.  
XX  
PR 23-JUL-1999; 99NZ-0336906.  
XX  
PA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.  
XX  
PI Scotti PD, Dearing SC, Greenwood DR, Newcomb RD:  
XX WPI: 2000-452375/39.  
XX DR P-PSDB; AA93750.  
XX



```
FT      /*tag= b
XX      MO200039165-A1.
XX      PD
XX      06-JUL-2000.
XX      PE
XX      23-DEC-1999; 99MO-N200227.
XX      PR
XX      23-DEC-1998; 98NZ-0333568.
XX      PR
XX      23-JUL-1999; 99NZ-0336906.
XX      (HOKT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
XX      PI
XX      Scotti PD, Dearing SC, Greenwood DR, Newcomb RD;
XX      WPI; 2000-452375/39.
XX      DR
XX      P-PSDB: AAT93750.
XX      CC
XX      New Perna canaliculus serine protease inhibitor protein exhibiting
XX      PT anti-thrombin activity and divalent metal cation binding activity,
XX      PT useful as an anticoagulant agent and as a dietary supplement -
XX      PS
XX      Claim 11: Page 10-11; 44pp; English.
XX      CC
XX      The present sequence encodes a serine protease inhibitor
XX      CC protein. The protein is isolated from the green-lipped mussel
XX      CC (Perna canaliculus), and exhibits, inter alia, anti-thrombin activity
XX      CC and divalent metal cation binding activity. The serine protease
XX      CC inhibitor protein has a molecular weight of about 55 kilo Daltons.
XX      CC The protein, and its fragments, are useful in medicaments, in food,
XX      CC as dietary supplements or as bioremediation agents. In the dietary
XX      CC supplements, the protein is associated with or bound to at least one
XX      CC divalent cation (such as calcium, magnesium or zinc) of dietary
XX      CC significance. The proteins or their fragments are also useful as
XX      CC anticoagulant agents.
XX      CC
XX      SQ Sequence 1611 BP; 499 A; 348 C; 360 G; 402 T; 2 other:
XX
XX      Query Match 100.0%; Score 1491; DB 21; Length 1611;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY 1 GATGGGGACATGATGAGTGGGAGAGAAAGATGACCATGACGACCAACGAT 60
XX      Db 1 GATGGGGACATGATGAGTGGGAGAGAAAGATGACCATGACGACCAACGAT 60
XX      QY 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGAT 120
XX      Db 61 GATCACCATGACGACCATGATGATGATGATGATGATGATGATGATGATGAT 120
XX      QY 121 GAACCAAAACCTTCATATGAGTGGTACAGCCTTCACCATGATGATGATGAT 180
XX      Db 121 GAACCAAAACCTTCATATGAGTGGTACAGCCTTCACCATGATGATGATGAT 180
XX      QY 181 TCACAGAGAGGTCATGAGAGCTGTTTATCTAGAACCTTCATGATGATGAT 240
XX      Db 181 TCACAGAGAGGTCATGAGAGCTGTTTATCTAGAACCTTCATGATGATGAT 240
XX      QY 241 GAAGACCATGACGACCATCATCATGATGATGATGATGATGATGATGATGAT 300
XX      Db 241 GAAGACCATGACGACCATCATCATGATGATGATGATGATGATGATGATGAT 300
XX      QY 301 GATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
XX      Db 301 GATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
XX      QY 361 GACCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
XX      Db 361 GACCTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
XX      QY 421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
XX      Db 421 TGGTGGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
```

```
QY 481 ATTTTACAGGAGTACACACCGATGCTGATACCCAGCAGTATGATGCTGTGTGTT 540
Db 481 ATTTTACAGGAGTACACACCGATGCTGATACCCAGCAGTATGATGCTGTGTGTT 540
QY 541 ATTTGCTATGGAAGAGTCTGGCCAGAAACAGAGAGTCTGATACATGACGATGAGGAA 600
Db 541 ATTTGCTATGGAAGAGTCTGGCCAGAAACAGAGAGTCTGATACATGACGATGAGGAA 600
QY 601 GATTAACCTGACATTTATGCTGATGATGATGATGATGATGATGATGATGATGAT 660
Db 601 GATTAACCTGACATTTATGCTGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 CTTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 661 CTTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 GTGTCCTACCATTTAGAGGATTTAATGTAAGTATGATGATGATGATGATGATGAT 780
Db 721 GTGTCCTACCATTTAGAGGATTTAATGTAAGTATGATGATGATGATGATGATGAT 780
QY 781 GTACAGATCTACGCCAACGCGTGACCTGACAGTATGATGATGATGATGATGATGAT 840
Db 781 GTACAGATCTACGCCAACGCGTGACCTGACAGTATGATGATGATGATGATGATGAT 840
QY 841 GATCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 841 GATCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 CATGAGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 901 CATGAGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 GTCTGGGAGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 961 GTCTGGGAGCTGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 GCTGTTGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db 1021 GCTGTTGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 GTTGTGACAGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 GTTGTGACAGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 TTGCAACAGACCCCTGGAGGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1141 TTGCAACAGACCCCTGGAGGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 AGTGAAGACTTGTACATCATGATGATGATGATGATGATGATGATGATGATGAT 1260
Db 1201 AGTGAAGACTTGTACATCATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 CATGCTGTGCTACTTCCTTAGGACAGATGATGATGATGATGATGATGATGAT 1320
Db 1261 CATGCTGTGCTACTTCCTTAGGACAGATGATGATGATGATGATGATGATGAT 1320
QY 1321 AGACCTGGTACCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1380
Db 1321 AGACCTGGTACCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 ACCTTGTATCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
Db 1381 ACCTTGTATCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 GGACATGAGGTGAGAGTGAAGAGGTTGCTTCTGTTTATGAGACGGGCA 1491
Db 1441 GGACATGAGGTGAGAGTGAAGAGGTTGCTTCTGTTTATGAGACGGGCA 1491
```

RESULT 3  
ABL06474  
ID ABL06474 standard: cDNA: 3211 BP.

```
XX ABL06474;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13904.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KM
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PE
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB62371.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
CC
XX Claim 1; SEQ ID NO 13904; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3211 BP; 922 A; 708 C; 688 G; 893 T; 0 other;
SQ
Query Match 1.7%; Score 26; DB 23; Length 3211;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CACAGATATTCCTGGATCAATATCTT 947
DB 2629 CACAGATATTCCTGGATCAATATCTT 2654
RESULT 4
ABL0542
ID ABL0542 standard; cDNA; 6398 BP.
AC
XX
XX ABL0542;
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 26108.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
```

```
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PE
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB6439.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
CC
XX Claim 1; SEQ ID NO 26108; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6398 BP; 2021 A; 1289 C; 1522 G; 1566 T; 0 other;
SQ
Query Match 1.7%; Score 26; DB 23; Length 6398;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 922 CACAGATATTCCTGGATCAATATCTT 947
DB 665 CACAGATATTCCTGGATCAATATCTT 690
RESULT 5
ABL23878
ID ABL23878 standard; DNA; 2171 BP.
AC
XX
XX ABL23878;
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23107.
DE
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
KM pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PE
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
```

|          |  |
|----------|--|
| XX       | New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -  |
| PT       | Interactions -   |
| PS       | Claim 1; SEQ ID NO 23107; 21pp + Sequence Listing; English.  |
| CC       | The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA sequences (ABB57737-ABB72072).   |
| CC       | (ABB57737-ABB72072).   |
| CC       | The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.  |
| CC       |  |
| CC       |  |
| SO       | Sequence 2171 BP; 589 A; 498 C; 475 G; 609 T; 0 other;   |
| OY       | Query Match<br>Best Local Similarity 100.0%; Score 25; DB 23; Length 2171;<br>Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |
| DB       | 68 ATGACGACCATGATGCATGCATGA 92<br>     <br>228 ATGACGACCATGATGCATGCATGA 252  |
| RESULT 6 |  |
| ID       | AAC39999/C   |
| XX       | AAC39999 standard; DNM; 1599 BP.   |
| XX       | AAC39999;  |
| DX       |  |
| DT       | 17-OCT-2000 (first entry)  |
| DE       | Arabidopsis thaliana DNA fragment SEQ ID NO: 26678.  |
| XX       |  |
| KW       | Hybridisation assay; genetic mapping; gene expression control;<br>protein identification; signal transduction pathway;<br>metabolic pathway; promoter; termination sequence; ss.   |
| XX       |  |
| OS       | Arabidopsis thaliana.  |
| PN       | EP1033405-A2.  |
| PD       |  |
| FD       | 06-SEP-2000.   |
| XX       |  |
| PX       | .25-FEB-2000; 2000EP-0301439.  |
| XX       |  |
| PR       | 25-FEB-1999; 99US-0121825.<br>05-MAR-1999; 99US-0123180.<br>09-MAR-1999; 99US-0123548.<br>23-MAR-1999; 99US-0125788.<br>25-MAR-1999; 99US-0126264.<br>29-MAR-1999; 99US-0126785.<br>01-APR-1999; 99US-0127462.<br>06-APR-1999; 99US-0128234.<br>08-APR-1999; 99US-0128714.<br>16-APR-1999; 99US-0128645.<br>19-APR-1999; 99US-0130077.<br>21-APR-1999; 99US-0130449.<br>23-APR-1999; 99US-0130510.<br>23-APR-1999; 99US-0130891.<br>28-APR-1999; 99US-0131449.<br>30-APR-1999; 99US-0132407.<br>04-MAY-1999; 99US-0132484.<br>05-MAY-1999; 99US-0132485.<br>06-MAY-1999; 99US-0132486.<br>06-MAY-1999; 99US-0132487. |

|    |             |               |
|----|-------------|---------------|
| PR | 07-MAY-1999 | 9905-0138266  |
| PR | 11-MAY-1999 | 9905-01384256 |
| PR | 14-MAY-1999 | 9905-01384219 |
| PR | 14-MAY-1999 | 9905-01384218 |
| PR | 14-MAY-1999 | 9905-01384219 |
| PR | 14-MAY-1999 | 9905-01384221 |
| PR | 16-MAY-1999 | 9905-01384370 |
| PR | 16-MAY-1999 | 9905-01384376 |
| PR | 18-MAY-1999 | 9905-01384941 |
| PR | 21-MAY-1999 | 9905-01385353 |
| PR | 21-MAY-1999 | 9905-01385629 |
| PR | 25-MAY-1999 | 9905-01386021 |
| PR | 27-MAY-1999 | 9905-01386392 |
| PR | 28-MAY-1999 | 9905-01387282 |
| PR | 01-JUN-1999 | 9905-01387222 |
| PR | 03-JUN-1999 | 9905-01387528 |
| PR | 04-JUN-1999 | 9905-01387502 |
| PR | 07-JUN-1999 | 9905-01387724 |
| PR | 08-JUN-1999 | 9905-01388094 |
| PR | 10-JUN-1999 | 9905-01388540 |
| PR | 10-JUN-1999 | 9905-01388487 |
| PR | 14-JUN-1999 | 9905-01391149 |
| PR | 16-JUN-1999 | 9905-01394552 |
| PR | 16-JUN-1999 | 9905-01394592 |
| PR | 17-JUN-1999 | 9905-01394932 |
| PR | 18-JUN-1999 | 9905-01394554 |
| PR | 18-JUN-1999 | 9905-01394655 |
| PR | 18-JUN-1999 | 9905-01394657 |
| PR | 18-JUN-1999 | 9905-01394657 |
| PR | 18-JUN-1999 | 9905-01394657 |
| PR | 18-JUN-1999 | 9905-01394659 |
| PR | 18-JUN-1999 | 9905-01394661 |
| PR | 18-JUN-1999 | 9905-01394661 |
| PR | 18-JUN-1999 | 9905-01394663 |
| PR | 18-JUN-1999 | 9905-01394663 |
| PR | 18-JUN-1999 | 9905-01397673 |
| PR | 21-JUN-1999 | 9905-01398177 |
| PR | 22-JUN-1999 | 9905-01403839 |
| PR | 23-JUN-1999 | 9905-01403583 |
| PR | 24-JUN-1999 | 9905-01403534 |
| PR | 24-JUN-1999 | 9905-01406655 |
| PR | 28-JUN-1999 | 9905-01408283 |
| PR | 29-JUN-1999 | 9905-01409911 |
| PR | 30-JUN-1999 | 9905-01412827 |
| PR | 01-JUL-1999 | 9905-01414842 |
| PR | 01-JUL-1999 | 9905-01421554 |
| PR | 02-JUL-1999 | 9905-01420255 |
| PR | 06-JUL-1999 | 9905-01423800 |
| PR | 08-JUL-1999 | 9905-01423803 |
| PR | 09-JUL-1999 | 9905-01423920 |
| PR | 12-JUL-1999 | 9905-01429277 |
| PR | 13-JUL-1999 | 9905-01433654 |
| PR | 14-JUL-1999 | 9905-01433654 |
| PR | 15-JUL-1999 | 9905-01440005 |
| PR | 16-JUL-1999 | 9905-01440005 |
| PR | 16-JUL-1999 | 9905-01440006 |
| PR | 19-JUL-1999 | 9905-01444351 |
| PR | 19-JUL-1999 | 9905-01444351 |
| PR | 19-JUL-1999 | 9905-01444351 |
| PR | 19-JUL-1999 | 9905-01444332 |
| PR | 19-JUL-1999 | 9905-01444333 |
| PR | 19-JUL-1999 | 9905-01444334 |
| PR | 19-JUL-1999 | 9905-01444335 |
| PR | 19-JUL-1999 | 9905-01445086 |
| PR | 21-JUL-1999 | 9905-01450887 |
| PR | 22-JUL-1999 | 9905-01450887 |
| PR | 22-JUL-1999 | 9905-01450887 |
| PR | 22-JUL-1999 | 9905-01450889 |
| PR | 22-JUL-1999 | 9905-01451592 |

PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148655.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149923.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150684.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151308.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159325.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 1.4%; Score 21; DB 21; Length 1599;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 77 ATGCATGCATGCATGAACAA 97  
DB 139 ATGCATGCATGCATGAACAA 119  
|||||  
  
RESULT 7  
ID ABL15963 standard; cDNA; 3309 BP.  
XC ABL15963;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 42371.  
DE Drosophila: developmental biology; cell signalling; insecticide;  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
KW Drosophila melanogaster.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
XX Ventler JC, Adams M, Li PMD, Myers EW;  
PI MPI: 2001-656860/75.  
XX P-PSDB; ABB71860.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1: SEQ ID NO 42371; 21bp + Sequence Listing: English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 3309 BP: 897 A; 893 C; 874 G; 645 T; 0 other;



Query Match 1.4%; Score 21; DB 23; Length 3309;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 GACGACGATGACGACGAC 57  
 ||||||||||||||||||  
 DB 2674 GACGACGATGACGACGAC 2694

## RESULT 8

AA118924/c  
 ID AA118924 standard; DNA; 4909 BP.

AC AA118924;

DT 17-JAN-1997 (first entry)

DE Plasmid pA1261.

XX Spider; dragline protein; variant; monomer; polymer; circular;

KW fibre forming region; Spidroin 1; Nephrila clavipes; DPL; mimic;

KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;

KW rope; surgical suture; implant; reinforcement; film; coating; ss.

OS Synthetic.

PN MO9429450-A2.

PD 22-DEC-1994.

PF 15-JUN-1994; 94WO-US06689.

PR 15-JUN-1993; 93US-0077600.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Fahnstock SR;

DR WPI; 1995-036479/05.

XX Example 1; Fig 13; 168pp; English.

This sequence represents the complete nucleotide sequence of the plasmid pA1261. This plasmid was used in the construction of the vector pFP510 which was used to express synthetic spider dragline variants, DP-1A.9 and DP-1B.9. pA1261 comprises a replication origin active in E. coli, a selectable genetic marker which is a gene conferring resistance to ampicillin, sites for the restriction endonucleases BamHI and BglII with no essential sequences between them, and a third restriction site for PstI located within the selectable marker which produces cohesive ends incompatible with those produced by BamHI and BglII. The polypeptide monomers are variants based on a consensus sequence derived from the fibre forming regions of spider dragline proteins, esp. the natural protein 1 (Spidroin 1) from Nephrila clavipes. Synthetic analogues of DP1 were designed to mimic the repeating consensus sequence of the natural protein and the pattern of variation among individual repeats. DP-1A analogues are composed of a tandemly repeated 101 amino acid monomer which comprises four repeats which differ from the consensus sequence given in AAM06201, according to the pattern (1)-(5):  
 (1) the poly-alanine sequence varies in length from 0-7 residues; (2) when the entire poly-alanine sequence is deleted, so also is the surrounding sequence encompassing AGGAGGAGGAGG; (3) aside from the poly-alanine sequence, deletions usually encompass integral multiples of three consecutive residues; (4) deletion of GYG is generally accompanied by deletion of GRG in the same sequence; and  
 (5) a repeat in which the entire poly-alanine sequence is deleted is generally preceded by a repeat containing six alanine

CC residues.  
 CC The proteins may be used to produce fibres of high tensile strength and  
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials  
 CC for implants, plastic reinforcements, films, coatings, etc.  
 CC  
 XX

SO Sequence 4909 BP; 1216 A; 1251 C; 1258 G; 1164 T; 0 other;

Query Match 1.4%; Score 21; DB 16; Length 4909;  
 Best Local Similarity 100.0%; Pred. No. 4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1332 CCTTGGTGTATATAGATGA 1352  
 ||||||||||||||||||  
 DB 2163 CCTTGGTGTATATAGATGA 2143

## RESULT 9

ABL15962/c  
 ID ABL15962 standard; cDNA; 6699 BP.

AC ABL15962;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide seq ID NO 42368.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

XX Venter JC, Adams M, LI PWD, Myers EW;

PI WPI; 2001-656860/75.

DR P-PSDB; ABB71859.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Claim 1; SEQ ID NO 42368; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention includes genomic DNA sequences (ABL16176-ABJ30511); expressed DNA sequences (ABB57737-ABB72072).  
 The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/publshd\_pcl\_sequences.

CC Sequence 6699 BP; 1663 A; 1551 C; 1541 G; 1944 T; 0 other;

Query Match 1.4%; Score 21; DB 23; Length 6699;  
 Best Local Similarity 100.0%; Pred. No. 4.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 37 GACGACGATGACGACGAC 57  
 ||||||||||||||||||

DB 1636 GACCACCATGACGACCAC 1616

## RESULT 10

ABLS1954 standard; DNA; 40023 BP.

AC ABLS1954;

DT 11-JUL-2002 (first entry)

XX Human solute carrier family 18 member 2 (SLC18A2) gene SEQ ID NO:1.

XX Human: solute carrier family 18 member 2; SLC18A2; vesicular monoamine;

KV vesicular monoamine transporter; VMAT2; polymorphic site; SNP;

KV single nucleotide polymorphism; anti-inflammatory; neuroleptic;

KV haplotyping; genotyping; respiratory inflammatory disease;

XX neuropsychiatric disorder; monoenergetic brain system; gene; ds.

OS Homo sapiens.

XX Key

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

FT misc\_feature

29050 /tag- T

29230 /note- "PS18: polymorphic base G or T"

29230 /tag- S

29395 /note- "PS19: polymorphic base C or T"

29395 /tag- T

32343 /note- "PS20: polymorphic base C or T"

32572 /tag- U

32572 /note- "PS21: polymorphic base C or T"

39464 /tag- V

39464 /note- "PS22: polymorphic base A or G"

39766 /tag- W

39848 /note- "PS23: polymorphic base T or C"

39848 /tag- X

39910 /note- "PS24: polymorphic base C or T"

39910 /tag- Y

39984 /note- "PS25: polymorphic base A or C"

39984 /tag- Z

39984 /note- "PS26: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

39984 /tag- aa

39984 /note- "PS27: polymorphic base T or C"

Sequence 40023 BP; 10020 A; 9128 C; 9557 G; 11291 T; 27 other;

Query Match 1.4%; Score 21; DB 24; Length 40023;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 GACGACCATGATGATGAT 90  
|||||  
DB 19958 GACGACCATGATGATGAT 19978

RESULT 11  
AAH45894/C  
ID AAH45894 standard; DNA; 65 BP.  
XX  
AC AAH45894;  
XX  
DT 17-SEP-2001 (first entry)  
XX  
XX C parvum coding sequence PCR primer JCA302.  
XX  
XX Cryptosporidium parvum infection; enteric pathogen; vaccine; cattle;  
KM epitope; PCR primer; ss.  
XX  
XX Cryptosporidium parvum.  
OS  
XX MO200145735-A2.  
PN  
XX 28-JUN-2001.  
PD  
XX 20-DEC-2000; 2000WO-EP13387.  
PF  
XX 21-DEC-1999; 99US-0171399.  
PR  
XX (MERI-) MERIAL.  
XX  
XX Audonnet J, Gallo G;  
PI  
XX WPI: 2001-417978/44.  
DR  
XX  
XX New immunological or vaccine composition for prevention, treatment and  
PT control of Cryptosporidium parvum and/or enteric infections in bovines,  
PT canines, has antigens from Cryptosporidium and another enteric pathogen  
PT  
XX  
XX Example 3; Page 34; 66pp; English.  
PS  
XX The present invention relates to a vaccine against Cryptosporidium  
CC parvum, which causes enteric infections in cattle, comprising two  
CC epitopes or antigens from the microorganism. This can also be used to  
CC treat enteric infections in canines, felines and equines. The present  
CC sequence is a PCR primer used to obtain the coding sequence of an epitope  
CC useful in the composition of the invention.  
CC  
XX  
SQ Sequence 65 BP; 9 A; 14 C; 14 G; 28 T; 0 other;

Query Match 1.3%; Score 20; DB 22; Length 65;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 GACCATGATGATGATGATGA 92  
|||||  
DB 42 GACCATGATGATGATGATGA 23

RESULT 12  
ABL93659  
ID ABL93659 standard; CDNA; 461 BP.  
XX  
AC ABL93659;  
XX  
XX 10-JUN-2002 (first entry)  
DT  
XX Arabidopsis thaliana nucleic acid sequence Ref:2027424 SEQ ID NO:424.

XX Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;  
KM genetic modification; gene; ss.  
XX  
XX Arabidopsis thaliana.  
OS  
XX US2002023280-A1.  
PN  
XX  
XX 21-FEB-2002.  
PD  
XX 26-JAN-2001; 2001US-0770444.  
PF  
XX 27-JAN-2000; 2000US-178502P.  
PR  
XX (GORL/) GORLACH J.  
PA (ANY/) AN Y.  
PA (HAM/) HAMILTON C M.  
PA (PRIC/) PRICE J L.  
PA (RAIN/) RAINES T M.  
PA (YUY/) YU Y.  
PA (RAME/) RAMEKA J G.  
PA (PAGE/) PAGE A.  
PA (MATH/) MATHEW A V.  
PA (LEDF/) LEDFORD B L.  
PA (WOES/) WOESSNER J P.  
PA (HAAS/) HAAS W D.  
PA (GARC/) GARCIA C A.  
PA (KRIC/) KRICKER M.  
PA (SLAT/) SLATER T.  
PA (DAVI/) DAVIS K R.  
PA (ALEE/) ALLEN K.  
PA (HOEF/) HOFFMAN N.  
PA (HURB/) HURBAN P.  
XX  
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;  
PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;  
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;  
PI Hurban P;  
XX  
XX WPI: 2002-267486/31.  
DR  
XX  
XX New Arabidopsis thaliana nucleic acid, for identifying homologous  
PT genes, producing compositions that modulate the expression or function  
PT of its encoded protein, and mapping functional regions of a protein -  
PT  
XX  
XX Claim 1; SEQ ID 424; 44pp; English.  
PS  
XX The present invention describes an Arabidopsis thaliana nucleic acid (I)  
CC comprising a sequence capable of hybridising under stringent conditions  
CC to a sequence (SI) selected from any one of the 999 sequences given in  
CC AB193236 to AB194234 (I) have insecticide and fungicide activities, and  
CC AB193236 to AB194234 (I) have insecticide and fungicide activities, and  
CC identifying homologous or related genes, in producing compositions that  
CC modulate the expression or function of their encoded proteins, mapping  
CC functional regions of the proteins, and in studying associated  
CC physiological pathways. (I) can also be used: (1) for the genetic  
CC manipulation of cells, particularly plant cells; (2) in screening assays  
CC of various plant strains to determine the strains that are best capable  
CC of withstanding or inhibiting production of a biosynthetic product in a plant;  
CC (4) as probes in mapping and in diagnosis, in genetic modification and  
CC for screening purposes, to generate additional copies of the nucleic  
CC acids, to generate ribozymes or antisense oligonucleotides, and as  
CC single-stranded DNA probes or as triple-strand forming oligonucleotides;  
CC and (5) for generating genetically modified transgenic organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
XX  
SQ Sequence 461 BP; 147 A; 77 C; 122 G; 115 T; 0 other;

Query Match 1.3%; Score 20; DB 24; Length 461;  
Best Local Similarity 100.0%; Pred. No. 12;







```

XX DE DNA encoding novel human diagnostic protein #3301.
XX
XX KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG03310.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensic; gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1: SEQ ID No 3301; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensic; gene mapping, identification of mutations
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197/AAS94364 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at http://wipo.int/pub/published\_pct\_sequences.
XX
XX SO Sequence 879 BP; 227 A; 196 C; 225 G; 231 T; 0 other;

Query Match 1.3%; Score 20; DB 23; Length 879;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 ATGATGATGATGATGAACA 96
Db 655 ATGATGATGATGATGAACA 674

```

Search completed: November 29, 2002, 10:09:02  
 Job time : 368 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 08:54:48 (Search time 60 Seconds  
(without alignments)  
7620.918 Million cell updates/sec

Title: US-09-868-760-6

Sequence: 1 gctgtgttataagagcgagca 1491

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size: 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents NA: \*  
1: /cgn2.6/prodata/1/lna/5A.COMB.seq: \*  
2: /cgn2.6/prodata/1/lna/5B.COMB.seq: \*  
3: /cgn2.6/prodata/1/lna/6A.COMB.seq: \*  
4: /cgn2.6/prodata/1/lna/6B.COMB.seq: \*  
5: /cgn2.6/prodata/1/lna/PCNUS.COMB.seq: \*  
6: /cgn2.6/prodata/1/lna/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID               | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 21    | 1.4         | 4909   | 4 US-08-556-978B-78 | Sequence 78, Appl |
| 2          | 19    | 1.3         | 53     | 2 US-08-687-865A-10 | Sequence 10, Appl |
| 3          | 19    | 1.3         | 53     | 4 US-09-043-711-10  | Sequence 10, Appl |
| 4          | 19    | 1.3         | 55     | 2 US-08-687-865A-11 | Sequence 11, Appl |
| 5          | 19    | 1.3         | 55     | 4 US-09-043-711-11  | Sequence 11, Appl |
| 6          | 19    | 1.3         | 96     | 3 US-08-737-336-2   | Sequence 23, Appl |
| 7          | 19    | 1.3         | 105    | 2 US-08-687-865A-23 | Sequence 23, Appl |
| 8          | 19    | 1.3         | 105    | 4 US-09-043-711-23  | Sequence 23, Appl |
| 9          | 19    | 1.3         | 129    | 2 US-08-679-865-38  | Sequence 38, Appl |
| 10         | 19    | 1.3         | 129    | 4 US-09-043-711-38  | Sequence 38, Appl |
| 11         | 19    | 1.3         | 133    | 2 US-08-680-876-38  | Sequence 38, Appl |
| 12         | 19    | 1.3         | 133    | 4 US-09-263-975-38  | Sequence 38, Appl |
| 13         | 19    | 1.3         | 153    | 2 US-08-680-876-41  | Sequence 41, Appl |
| 14         | 19    | 1.3         | 153    | 4 US-09-263-975-41  | Sequence 41, Appl |
| 15         | 19    | 1.3         | 234    | 2 US-08-687-865A-20 | Sequence 20, Appl |
| 16         | 19    | 1.3         | 234    | 4 US-09-043-711-20  | Sequence 20, Appl |
| 17         | 19    | 1.3         | 466    | 4 US-08-457-245-20  | Sequence 20, Appl |
| 18         | 19    | 1.3         | 782    | 4 US-09-004-731-61  | Sequence 91, Appl |
| 19         | 19    | 1.3         | 782    | 4 US-08-749-699-61  | Sequence 91, Appl |
| 20         | 19    | 1.3         | 782    | 4 US-09-004-729-91  | Sequence 91, Appl |
| 21         | 19    | 1.3         | 783    | 4 US-09-004-731-93  | Sequence 93, Appl |
| 22         | 19    | 1.3         | 783    | 4 US-08-749-699-93  | Sequence 93, Appl |
| 23         | 19    | 1.3         | 811    | 4 US-09-004-729-93  | Sequence 93, Appl |
| 24         | 19    | 1.3         | 811    | 4 US-09-004-731-88  | Sequence 88, Appl |
| 25         | 19    | 1.3         | 811    | 4 US-08-749-699-88  | Sequence 88, Appl |
| 26         | 19    | 1.3         | 831    | 4 US-09-004-729-88  | Sequence 88, Appl |
| 27         | 19    | 1.3         | 831    | 4 US-09-004-731-90  | Sequence 90, Appl |

|    |    |     |      |                     |                   |
|----|----|-----|------|---------------------|-------------------|
| 28 | 19 | 1.3 | 823  | 4 US-08-749-699-90  | Sequence 90, Appl |
| 29 | 19 | 1.3 | 823  | 4 US-09-004-729-90  | Sequence 90, Appl |
| 30 | 19 | 1.3 | 1056 | 2 US-08-701-191A-5  | Sequence 25, Appl |
| 31 | 19 | 1.3 | 1395 | 1 US-07-991-867B-25 | Sequence 25, Appl |
| 32 | 19 | 1.3 | 1395 | 1 US-08-107-755A-25 | Sequence 25, Appl |
| 33 | 19 | 1.3 | 1395 | 2 US-08-544-332-25  | Sequence 25, Appl |
| 34 | 19 | 1.3 | 1395 | 4 US-09-370-861A-25 | Sequence 28, Appl |
| 35 | 19 | 1.3 | 1478 | 4 US-09-545-814-28  | Sequence 30, Appl |
| 36 | 19 | 1.3 | 1478 | 4 US-09-545-814-30  | Sequence 30, Appl |
| 37 | 19 | 1.3 | 1599 | 3 US-08-737-336-5   | Sequence 7, Appl  |
| 38 | 19 | 1.3 | 1866 | 1 US-08-605-541B-11 | Sequence 11, Appl |
| 39 | 19 | 1.3 | 2081 | 2 US-09-096-982-7   | Sequence 7, Appl  |
| 40 | 19 | 1.3 | 2081 | 2 US-08-653-650A-7  | Sequence 7, Appl  |
| 41 | 19 | 1.3 | 4770 | 4 US-09-000-094-45  | Sequence 45, Appl |
| 42 | 19 | 1.3 | 6768 | 1 US-08-107-755A-1  | Sequence 1, Appl  |
| 43 | 19 | 1.3 | 8457 | 1 US-07-991-867B-1  | Sequence 1, Appl  |
| 44 | 19 | 1.3 | 8457 | 2 US-08-544-332-1   | Sequence 1, Appl  |
| 45 | 19 | 1.3 | 8457 | 4 US-09-370-861A-1  | Sequence 1, Appl  |

## ALIGNMENTS

RESULT 1  
US-08-556-978B-78/c  
Sequence 78, Application US/08556978B  
Patent No. 6268169  
GENERAL INFORMATION:  
APPLICANT: FAHNESTOCK, STEPHEN F.  
TITLE OF INVENTION: NOVEL RECOMBINANTLY PRODUCED  
NUMBER OF INVENTIONS: SPIDER SILK ANALOGS  
TITLE OF INVENTION: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD FOR WINDOWS 95  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/556,978B  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/077,600  
FILING DATE: JUNE 15, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: FLOYD, LINDA AXAMETHY  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: CR-9389-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-892-8112  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 78:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4909 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
US-08-556-978B-78  
Query Match 1.4%, Score 21, DB 4, Length 4909;  
Best Local Similarity 100.0%, Pred. No. 0.67;  
Matches 21, Conservative 0, Mismatches 0, Indels 0, Gaps 0;  
QY 1332 CCTGTGTATGTTATAGATGA 1352  
|||||

Db 2163 CCTGGTGTATGATGATGA 2143

## RESULT 2

US-08-687-865A-10/C

Sequence 10, Application US/08687865A

Patent No. 5955596

GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.

APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus

TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,865A

FILING DATE: 26-JUL-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Gordon, Alan M.

REGISTRATION NUMBER: 30,637

REFERENCE/DOCKET NUMBER: 33,250-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-687-865A-10

Query Match

Best Local Similarity 1.3%; Score 19; DB 2; Length 53;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92

Db 49 ACCATGATGATGATGATGA 31

RESULT 3

US-09-043-711-10/C

Sequence 10, Application US/09043711

Patent No. 6221365

GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.

APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus

TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/043,711

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/687,865

FILING DATE: 26-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gordon, Alan M.

REGISTRATION NUMBER: 30,637

REFERENCE/DOCKET NUMBER: 33,250-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-043-711-10

Query Match

Best Local Similarity 1.3%; Score 19; DB 4; Length 53;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92

Db 49 ACCATGATGATGATGATGA 31

RESULT 4

US-08-687-865A-11

Sequence 11, Application US/08687865A

Patent No. 5955596

GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.

APPLICANT: Zagursky, Robert J.

TITLE OF INVENTION: The Nuca Protein of Haemophilus

TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,865A

FILING DATE: 26-JUL-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Gordon, Alan M.

REGISTRATION NUMBER: 30,637

REFERENCE/DOCKET NUMBER: 33,250-00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3244

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 55 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-687-865A-11

Query Match 1.3%; Score 19; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92  
Db 9 ACCATGATGATGATGATGA 27

RESULT 5  
US-09-043-711-11  
Sequence 11, Application US/09043711  
Patent No. 6221965

GENERAL INFORMATION:  
APPLICANT: Jones, Kevin F.  
APPLICANT: Zagursky, Robert J.  
TITLE OF INVENTION: The Nuca Protein of Haemophilus  
TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07470  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/043,711  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/687,865  
FILING DATE: 26-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gordon, Alan M.  
REGISTRATION NUMBER: 30,637  
REFERENCE/DOCKET NUMBER: 33,250-00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3244  
TELEFAX: 201-831-3305  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 55 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-043-711-11

Query Match 1.3%; Score 19; DB 4; Length 55;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92  
Db 9 ACCATGATGATGATGATGA 27

RESULT 6  
US-08-737-336-2/c

Sequence 2, Application US/08737336  
Patent No. 6013262  
GENERAL INFORMATION:  
APPLICANT: FRAZER, Ian  
APPLICANT: ZHOU, Jian  
TITLE OF INVENTION: RECOMBINANT PAPILLOMA VIRUS L1  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737,336  
FILING DATE: 16-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/AU95/00292  
FILING DATE: 17-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PM 5667  
FILING DATE: 17-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 65064/118  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 96 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-737-336-2

Query Match 1.3%; Score 19; DB 3; Length 96;  
Best Local Similarity 100.0%; Pred. No. 6.1;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92  
Db 33 ACCATGATGATGATGATGA 15

RESULT 7  
US-08-687-865A-23/c  
Sequence 23, Application US/08687865A  
Patent No. 5955596  
GENERAL INFORMATION:

APPLICANT: Jones, Kevin F.  
APPLICANT: Zagursky, Robert J.  
TITLE OF INVENTION: The Nuca Protein of Haemophilus  
TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07470  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,865A  
FILING DATE: 26-JUL-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Gordon, Alan M.  
REGISTRATION NUMBER: 30,637  
REFERENCE/DOCKET NUMBER: 33,250-00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3244  
TELEFAX: 201-831-3305  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs  
STRANDEDNESS: single  
TYPE: nucleic acid  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-687-865A-23

Query Match 1.3%; Score 19; DB 2; Length 105;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92  
DB 36 ACCATGATGATGATGATGA 18

RESULT 8  
US-09-043-711-23/C  
Sequence 23, Application US/09043711  
Patent No. 6221365  
GENERAL INFORMATION:  
APPLICANT: Jones, Kevin F.  
APPLICANT: Zagursky, Robert J.  
TITLE OF INVENTION: The NUCA Protein of Haemophilus  
TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: U.S.A.  
ZIP: 07470  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/043,711  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/687,865  
FILING DATE: 26-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gordon, Alan M.  
REGISTRATION NUMBER: 30,637  
REFERENCE/DOCKET NUMBER: 33,250-00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-831-3244  
TELEFAX: 201-831-3305  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 105 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-043-711-23

Query Match 1.3%; Score 19; DB 4; Length 105;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92  
DB 36 ACCATGATGATGATGATGA 18

RESULT 9  
US-08-679-865-38/C  
Sequence 38, Application US/08679865  
Patent No. 5912137  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
APPLICANT: Cubitt, Andrew B.  
TITLE OF INVENTION: Assays for Protein Kinases Using  
TITLE OF INVENTION: Fluorescent Protein Substrates  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/679,865  
FILING DATE: 16-JUL-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Storella, John S.  
REGISTRATION NUMBER: 32,944  
REFERENCE/DOCKET NUMBER: 023072-069000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 129 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 13...129  
US-08-679-865-38

Query Match 1.3%; Score 19; DB 2; Length 129;  
Best Local Similarity 100.0%; Pred. No. 6.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92  
DB 45 ACCATGATGATGATGATGA 27

RESULT 10  
US-08-680-876-38/C  
Sequence 38, Application US/08680876  
Patent No. 5925558

```

: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Cubitt, Andrew B.
: TITLE OF INVENTION: Assays for Protein Kinases Using
: TITLE OF INVENTION: Fluorescent Protein Substrates
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/680,876
: FILING DATE: 16-JUL-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Storella, John S.
: REGISTRATION NUMBER: 32,944
: REFERENCE/DOCKET NUMBER: 023072-069200
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 38:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 129 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..129
: US-08-680-876-38

Query Match 1.3%; Score 19; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 ACCATGATGATGATGATGA 92
Db 45 ACCATGATGATGATGATGA 27

RESULT 11
US-09-263-975-38/C
: Sequence 38, Application US/09263975
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Cubitt, Andrew B.
: TITLE OF INVENTION: Assays for Protein Kinases Using
: TITLE OF INVENTION: Fluorescent Protein Substrates
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
```

```

: APPLICATION NUMBER: US/09/263,975
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/679,865
: FILING DATE: 16-JUL-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Storella, John S.
: REGISTRATION NUMBER: 32,944
: REFERENCE/DOCKET NUMBER: 023072-069000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 38:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 129 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..129
: US-09-263-975-38

Query Match 1.3%; Score 19; DB 4; Length 129;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 ACCATGATGATGATGATGA 92
Db 45 ACCATGATGATGATGATGA 27

RESULT 12
US-08-679-865-41/C
: Sequence 41, Application US/08679865
: Patient No 5912137
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Cubitt, Andrew B.
: TITLE OF INVENTION: Assays for Protein Kinases Using
: TITLE OF INVENTION: Fluorescent Protein Substrates
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, Eighth Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/679,865
: FILING DATE: 16-JUL-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Storella, John S.
: REGISTRATION NUMBER: 32,944
: REFERENCE/DOCKET NUMBER: 023072-069000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 153 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
```

Mon Dec 2 12:19:18 2002

us-09-868-760-6.oli.rni

Page 6

```
; MOLECULE TYPE: DNA
; US-08-679-865-41
Query Match 1.3%; Score 19; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 ACCATGATGATGATGATGA 92
   |||
Db 45 ACCATGATGATGATGATGA 27

RESULT 13
US-08-680-876-41/C
; Sequence 41, Application US/08680876
; Patent No. 5925558
; GENERAL INFORMATION:
; APPLICANT: Tsiem, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,876
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 023072-069200
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-680-876-41

Query Match 1.3%; Score 19; DB 2; Length 153;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 ACCATGATGATGATGATGA 92
   |||
Db 45 ACCATGATGATGATGATGA 27

RESULT 14
US-09-263-975-41/C
; Sequence 41, Application US/09263975
; Patent No. 6248550
; GENERAL INFORMATION:
; APPLICANT: Tsiem, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/263,975
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/679,865
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 023072-069000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-09-263-975-41

Query Match 1.3%; Score 19; DB 4; Length 153;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 ACCATGATGATGATGATGA 92
   |||
Db 45 ACCATGATGATGATGATGA 27

RESULT 15
US-08-687-865A-20/C
; Sequence 20, Application US/08687865A
; Patent No. 5955596
; GENERAL INFORMATION:
; APPLICANT: Jones, Kevin F.
; APPLICANT: Zagursky, Robert J.
; APPLICANT: Ooi, Peggy
; TITLE OF INVENTION: The Nuca Protein of Haemophilus
; TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,865A
; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Gordon, Alan M.
```

```

; REGISTRATION NUMBER: 30,637
; REFERENCE/DOCKET NUMBER: 33,250-00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-831-3244
; TELEFAX: 201-831-3305
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 21..227
; US-08-687-865A-20

```

```

Query Match 1.3%; Score 19; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 74 ACCATGATGATGATGATGA 92
Db 53 ACCATGATGATGATGATGA 35

```

Search completed: November 29, 2002, 10:08:57  
Job time : 72 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 09:09:37 : Search time 67 Seconds

(without alignments)  
8570.256 Million cell updates/sec

Title: US-09-868-760-6

Perfect score: 1491  
Sequence: 1 gaggggagcagtgtaacga.....gctgtgtacagacggagca 1491

Scoring table: OLIGO\_MNC  
Gapop 60.0 , Gapext 60.0

Searched: 341543 seqs, 192557720 residues

Word size : 0

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications, NA:\*

- 1: /cgn2\_6/ptodata1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata1/pubpna/PCRT\_NEM\_PUB.seq:\*
- 3: /cgn2\_6/ptodata1/pubpna/US06\_NEM\_PUB.seq:\*
- 4: /cgn2\_6/ptodata1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata1/pubpna/US07\_NEM\_PUB.seq:\*
- 6: /cgn2\_6/ptodata1/pubpna/PCRTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata1/pubpna/US08\_NEM\_PUB.seq:\*
- 8: /cgn2\_6/ptodata1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata1/pubpna/US09\_NEM\_PUB.seq:\*
- 10: /cgn2\_6/ptodata1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata1/pubpna/US10\_NEM\_PUB.seq:\*
- 12: /cgn2\_6/ptodata1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata1/pubpna/US60\_NEM\_PUB.seq:\*
- 14: /cgn2\_6/ptodata1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                     | Description       |
|------------|-------|-------------|--------|------------------------|-------------------|
| 1          | 20    | 1.3         | 461    | 10 US-09-770-444-424   | Sequence 424, App |
| 2          | 19    | 1.3         | 1356   | 9 US-09-938-842A-1266  | Sequence 1266, Ap |
| 3          | 19    | 1.3         | 104    | 10 US-09-864-761-23649 | Sequence 23649, A |
| 4          | 13    | 1.3         | 104    | 10 US-09-864-761-24379 | Sequence 24379, A |
| 5          | 13    | 1.3         | 129    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 6          | 13    | 1.3         | 142    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 7          | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 8          | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 9          | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 10         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 11         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 12         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 13         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 14         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 15         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 16         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 17         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 18         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 19         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |
| 20         | 13    | 1.3         | 133    | 10 US-09-864-761-28873 | Sequence 28873, A |

|      |    |     |      |                        |                    |
|------|----|-----|------|------------------------|--------------------|
| c 20 | 18 | 1.2 | 266  | 10 US-09-878-574-14896 | Sequence 14896, A  |
| c 21 | 18 | 1.2 | 286  | 10 US-09-878-574-5372  | Sequence 5372, Ap  |
| c 22 | 18 | 1.2 | 451  | 10 US-09-770-444-667   | Sequence 667, App  |
| c 23 | 18 | 1.2 | 453  | 10 US-09-924-035A-231  | Sequence 231, App  |
| c 24 | 18 | 1.2 | 507  | 10 US-09-867-701-988   | Sequence 988, App  |
| c 25 | 18 | 1.2 | 511  | 10 US-09-864-761-8959  | Sequence 8959, Ap  |
| c 26 | 18 | 1.2 | 547  | 10 US-09-864-761-8835  | Sequence 8835, Ap  |
| c 27 | 18 | 1.2 | 708  | 9 US-09-938-842A-1875  | Sequence 1875, App |
| c 28 | 18 | 1.2 | 3072 | 10 US-09-841-739-3     | Sequence 3, App1   |
| c 29 | 18 | 1.2 | 3133 | 10 US-09-841-739-6     | Sequence 1, App1   |
| c 30 | 18 | 1.2 | 3612 | 10 US-09-841-739-4     | Sequence 4, App1   |
| c 31 | 18 | 1.2 | 3613 | 10 US-09-841-739-4     | Sequence 12, App1  |
| c 32 | 18 | 1.2 | 3613 | 10 US-09-841-739-4     | Sequence 6, App1   |
| c 33 | 18 | 1.2 | 6501 | 9 US-10-029-413A-3     | Sequence 3, App1   |
| c 34 | 18 | 1.2 | 7193 | 9 US-10-029-413A-17    | Sequence 17, App1  |
| c 35 | 18 | 1.2 | 7193 | 9 US-10-029-413A-19    | Sequence 19, App1  |
| c 36 | 18 | 1.2 | 7362 | 9 US-10-029-413A-15    | Sequence 15, App1  |
| c 37 | 18 | 1.2 | 7560 | 12 US-10-100-912-3     | Sequence 3, App1   |
| c 38 | 18 | 1.2 | 7746 | 10 US-09-778-927A-20   | Sequence 20, App1  |
| c 39 | 18 | 1.1 | 32   | 10 US-09-995-297-25    | Sequence 9, App1   |
| c 40 | 17 | 1.1 | 57   | 10 US-09-905-114-9     | Sequence 25, App1  |
| c 41 | 17 | 1.1 | 90   | 10 US-09-969-373-682   | Sequence 682, App  |
| c 42 | 17 | 1.1 | 146  | 10 US-09-864-761-27845 | Sequence 27845, A  |
| c 43 | 17 | 1.1 | 151  | 10 US-09-864-761-23373 | Sequence 23373, A  |
| c 44 | 17 | 1.1 | 164  | 10 US-09-864-761-24663 | Sequence 24663, A  |
| c 45 | 17 | 1.1 | 198  | 10 US-09-864-761-28856 | Sequence 28856, A  |

ALIGNMENTS

RESULT 1  
US-09-770-444-424  
Sequence 424, Application US/09770444  
Patent No. US2002023280A1  
GENERAL INFORMATION:  
APPLICANT: Goriach, Jörn  
APPLICANT: Han, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Ramezani, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Mathew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kricker, Maia  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurdan, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2027 (PARR-016PRV)  
CURRENT APPLICATION NUMBER: US/09/770,444  
PRIORITY FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,502  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 424  
LENGTH: 461  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-444-424  
Query Match 1.3%; Score 20; DB 10; Length 461;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 GACCATGATGATGATGATGA 92  
|||||  
DB 322 GACCATGATGATGATGATGA 341

RESULT 2  
US-09-938-842A-1266/C  
Sequence 1266, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Xun  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRIPI300-3  
CURRENT APPLICATION NUMBER: US/09/938, 842A  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 1266  
LENGTH: 1356  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1266

Query Match 1.3%; Score 20; DB 9; Length 1356;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 ATGATGATGATGATGAACA 96  
|||||  
DB 259 ATGATGATGATGATGAACA 240

RESULT 3  
US-09-864-761-23649/C  
Sequence 23649, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669

FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 23649  
LENGTH: 104  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC012619.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3  
OTHER INFORMATION: NT HIT: AF148532.1, EVALUATE 3.50e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BE963379.2, EVALUATE 1.30e+00  
US-09-864-761-23649

Query Match 1.3%; Score 19; DB 10; Length 104;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 ACCATGATGATGATGATGA 92  
|||||  
DB 74 ACCATGATGATGATGATGA 56

RESULT 4  
US-09-864-761-24379/C  
Sequence 24379, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669

```

; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24379
; LENGTH: 104
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC020919.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
; OTHER INFORMATION: NT HIT: AF148532.1, EVALUATE 3.50e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE963379.2, EVALUATE 1.30e+00
; US-09-864-761-24379

Query Match      1.3%  Score 19; DB 10; Length 104;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      74 ACCATGATGATGATGATGA 92
      |||||||||||||||
DB      74 ACCATGATGATGATGATGA 56

RESULT 5
US-09-864-681-38/C
; Sequence 38 Application US/09884681
; Patient No. US20020061546A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: Assays for Protein Kinases Using
;               Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/884,681
; FILING DATE: 19-Jun-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,865
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 13..129
; US-09-864-681-38

Query Match      1.3%  Score 19; DB 10; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      74 ACCATGATGATGATGATGA 92
      |||||||||||||||
DB      45 ACCATGATGATGATGATGA 27

RESULT 6
US-09-864-761-28873/C
; Sequence 28873, Application US/09864761
; Patient No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-28
; PRIOR APPLICATION NUMBER: US 09/633,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 28873  
LENGTH: 142  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC025644.2  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.59  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.76  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.57  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57  
US-09-864-761-28873

Query Match 1.3%; Score 19; DB 10; Length 142;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 ATGATGATGATGATGAAC 95  
Db 105 ATGATGATGATGATGAAC 87

RESULT 7  
US-09-884-681-41/C  
Sequence 41, Application US/09884681  
Patent No. US20020061546A1  
GENERAL INFORMATION:  
APPLICANT: Tsien, Roger Y.  
Cubitt, Andrew B.  
TITLE OF INVENTION: Assays for Protein Kinases Using  
Fluorescent Protein Substrates  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/884,681  
FILING DATE: 19-Jun-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/679,865  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Scirella, John S.  
REGISTRATION NUMBER: 32,944  
REFERENCE/DOCKET NUMBER: 023072-069000  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 153 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 41:  
US-09-884-681-41

Query Match 1.3%; Score 19; DB 10; Length 153;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 ACCATGATGATGATGA 92  
Db 45 ACCATGATGATGATGA 27

RESULT 8  
US-09-770-444-955  
Sequence 955, Application US/09770444  
Patent No. US20020023280A1  
GENERAL INFORMATION:  
APPLICANT: Gorlach, Jörn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Teng  
APPLICANT: Kameoka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Woessner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Kricker, Maya  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hurban, Patrick  
TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
FILE REFERENCE: 2027 (PARA-016PV)  
CURRENT APPLICATION NUMBER: US/09/770,444  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,502  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 955  
LENGTH: 439  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(439)  
OTHER INFORMATION: n = A,T,C or G  
US-09-770-444-955

Query Match 1.3%; Score 19; DB 10; Length 439;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 837 ATATGATCCTCATGAAGAT 855  
Db 373 ATATGATCCTCATGAAGAT 391

```
RESULT 9
US-09-864-761-6917/C
; Sequence 6917, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6917
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012619.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
US-09-864-761-6917

Query Match 1.3%; Score 19; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 412 ACCATGATGATGATGATCA 394

RESULT 10
US-09-864-761-7675/C
; Sequence 7675, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7675
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC020919.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
US-09-864-761-7675

Query Match 1.3%; Score 19; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 5.3;
```

|    |         |                     |              |    |            |    |        |    |      |    |
|----|---------|---------------------|--------------|----|------------|----|--------|----|------|----|
|    | Matches | 19,                 | Conservative | 0; | Mismatches | 0; | Indels | 0; | Gaps | 0; |
| OY | 74      | ACCATGATGATGATGATGA | 92           |    |            |    |        |    |      |    |
| Db | 412     | ACCATGATGATGATGATGA | 394          |    |            |    |        |    |      |    |

RESULT 11  
US-09-864-761-12293/c  
Sequence 12293, Application US/09864761

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```

CURRENT APPLICATION NUMBER: 05/09/664,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/653,366
PRIOR FILING DATE: 2000-09-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-11-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonex Sequence Listing Engine vers. 1.1.1
SEQ ID NO: 12293
LENGTH: 574

```

```

? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC025644.2
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.59
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.76
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.57
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
? OS: 09-864-761-12293

```

| Query Match | 1.38; | Score 19; | DB 10; | Length 574; |
|-------------|-------|-----------|--------|-------------|
|-------------|-------|-----------|--------|-------------|

|    |                        |        |         |                  |               |
|----|------------------------|--------|---------|------------------|---------------|
|    | Best Local Similarity  | 100.0% | Pred.   | No. 5.3          |               |
| Oy | 77 ATGATGATGATGATGAAC  | 95     | Matches | 19; Conservative | 0; Mismatches |
| Db | 198 ATGATGATGATGATGAAC | 180    |         |                  | Indels        |
|    |                        |        |         |                  | Gaps          |

```

QY      77  ATGATGATGATGATGAAC  95
          |||||
Db     198  ATGATGATGATGATGAAC  180

```

RESULT 12  
US-09-808-483-17

```

? Patent No. US20020001824A1
?
? GENERAL INFORMATION:
? Applicant: Bayer Aktiengesellschaft
? Title of Invention: Ligand-gated anion channels of insect
? File Reference: de A 34 397
? Current Application Number: US/09/808,483
? Current Filing Date: 2001-03-14
? Prior Application Number: DE 100 136 19.2
? Prior Filing Date: 2000-03-18
? Number of Seq ID Nos: 18
? Software: PatentIn Ver. 2.1
? Seq ID No 17
? Length: 1257
? Type: DNA
?
? ORGANISM: Drosophila melanogaster
?
? FEATURE:
? Name/Key: CDS
? Location: (1)..(1254)
? OS-09-808-483-17

```

|                       |                 |                |           |              |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match           | 1.38;           | Score 19;      | DB 10;    | Length 1257; |
| Best Local Similarity | 100.0%;         | Pred. NO. 5.5; |           |              |
| Matches 19;           | Conservative 0; | Mismatches 0;  | Indels 0; | Gaps 0.      |

|    |     |                      |     |
|----|-----|----------------------|-----|
| QY | 346 | CATGCTGACCCCTGGTGACC | 364 |
|    |     |                      |     |
| Db | 775 | CATGCTGACCCCTGGTGACC | 793 |

RESULT 13  
US-09-845-157-1/c

Patient No. US20020090618A1  
 GENERAL INFORMATION:  
 Applicant: Smith J  
 TITLE OF INVENTION: 'Thermostable Reverse Transcriptases and Uses Thereof'  
 FILE REFERENCE: 0942,5040001  
 CURRENT APPLICATION NUMBER: US/09/845,157  
 PRIOR FILING DATE: 2001-05-01  
 PRIOR APPLICATION NUMBER: US 60//207,196  
 PRIOR FILING DATE: 2000-05-26  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO. 1  
 LENGTH: 2151  
 TYPE: DNA  
 ORGANISM: M-MLV reverse transcriptase gene  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(2151)  
 US-09-845-157-1

```

Query Match Score 19; DB 10; Length 2151;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;
OY 74 ACCATGATGATGATGATGATGA 92
      |||||
DB 33 ACCATGATGATGATGATGATGA 15

```

## RESULT 14

```

US-09-965-703-74/C
; Sequence 74, Application US/09965703
; Patent No. US20020119521A1
; GENERAL INFORMATION:
; APPLICANT: Rohm and Haas Company
; APPLICANT: Paili, Subba Reddy
; APPLICANT: Kapitskaya, Marianna Zinovjevna
; APPLICANT: Cress, Dean Ervin
; TITLE OF INVENTION: No. US20020119521A1e1 Ecdysone Receptor-Based Inducible Gene Expr
; FILE REFERENCE: A01020B
; CURRENT APPLICATION NUMBER: US/09/965,703
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/191,355
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/269,799
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: PCT/US01/09030
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 74
; LENGTH: 3157
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-965-703-74

Query Match          1.3%; Score 19; DB 10; Length 3157;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 ACCATGATGATGATGATGA 92
|||||
DB 33 ACCATGATGATGATGATGA 15

RESULT 15
US-09-969-373-341/C
; Sequence 341, Application US/09969373
; Patent No. US2002013852A1
; GENERAL INFORMATION:
; APPLICANT: Eifertz, Roger J.
; APPLICANT: Hauge, Brian M.
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 341
; LENGTH: 129
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-341

Query Match          1.2%; Score 18; DB 10; Length 129;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CATGATGATGATGATGAA 93
|||||
DB 79 CATGATGATGATGATGAA 62

```

Search completed: November 29, 2002, 10:10:34  
 Job time : 88 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 29, 2002, 10:09:13 ; Search time 2057 Seconds

(Without alignments)  
11739.169 Million cell updates/sec

Title: US-09-868-760-6

Sequence: 1991  
1 gaggggagacgtacga.....gctgtatagagcggca 1491

Scoring table: OLIGO-MNC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vtl:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description         |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1          | 26    | 1.7         | 1101   | 17 | CNS00HD3 | AL073332 Drosophila |
| 2          | 25    | 1.7         | 529    | 14 | BQ118156 | BQ118156 EST603732  |
| 3          | 23    | 1.5         | 541    | 9  | AI724181 | AI724181 RH121_B_8  |
| 4          | 22    | 1.5         | 302    | 9  | AU267147 | AU267147 AU267147   |
| 5          | 22    | 1.5         | 303    | 9  | AU269373 | AU269373 AU269373   |
| 6          | 22    | 1.5         | 330    | 9  | AI931386 | AI931386 u169d11.y  |

|    |    |     |     |    |           |                              |
|----|----|-----|-----|----|-----------|------------------------------|
| 7  | 22 | 1.5 | 500 | 17 | A239345   | A239345 RPT-23-3             |
| 8  | 22 | 1.5 | 699 | 17 | B140875   | B140875 B140875              |
| 9  | 21 | 1.4 | 171 | 12 | BF759175  | BF759175 RC4-CT047           |
| 10 | 21 | 1.4 | 368 | 13 | BI702679  | BI702679 IT56M03.Y           |
| 11 | 21 | 1.4 | 387 | 11 | A0845314  | A0845314 LMAUFV1_1           |
| 12 | 21 | 1.4 | 461 | 10 | AW944020  | AW944020 LP03320.3           |
| 13 | 21 | 1.4 | 465 | 13 | BM159117  | BM159117 1427412.Y           |
| 14 | 21 | 1.4 | 493 | 13 | BI946460  | BI946460 B5T59.5             |
| 15 | 21 | 1.4 | 649 | 10 | AV610738  | AV610738 AV610738            |
| 16 | 21 | 1.4 | 655 | 14 | BP024062  | BP024062 BP024062            |
| 17 | 21 | 1.4 | 691 | 17 | AQ257939  | AQ257939 nbd00119M           |
| 18 | 21 | 1.4 | 766 | 10 | BE469502  | BE469502 IPHdK0189           |
| 19 | 20 | 1.3 | 225 | 10 | AW156710  | AW156710 se29608.Y           |
| 20 | 20 | 1.3 | 244 | 12 | BG311971  | BG311971 UI-R-CVO-           |
| 21 | 20 | 1.3 | 292 | 17 | AZ992197  | AZ992197 2M0276E16           |
| 22 | 20 | 1.3 | 307 | 14 | BQ980111  | BQ980111 OGE10C16.           |
| 23 | 20 | 1.3 | 333 | 10 | AV552114  | AV552114 AV552114            |
| 24 | 20 | 1.3 | 344 | 12 | BG511671  | BG511671 sad09c05.           |
| 25 | 20 | 1.3 | 345 | 17 | BH400144  | BH400144 AC-ND-156           |
| 26 | 20 | 1.3 | 355 | 14 | BQ986763  | BQ986763 OGF10G05.           |
| 27 | 20 | 1.3 | 371 | 13 | BI099490  | BI099490 IP1.42.H0           |
| 28 | 20 | 1.3 | 375 | 14 | BQ539271  | BQ539271 MEST614-B           |
| 29 | 20 | 1.3 | 377 | 12 | BF324534  | BF324534 su25b06.Y           |
| 30 | 20 | 1.3 | 389 | 9  | AU230857  | AU230857 AU230857            |
| 31 | 20 | 1.3 | 404 | 10 | BE523097  | BE523097 M423CSTM            |
| 32 | 20 | 1.3 | 410 | 12 | BF521537  | BF521537 EST458013           |
| 33 | 20 | 1.3 | 411 | 17 | A286838   | A286838 ZM0177022            |
| 34 | 20 | 1.3 | 412 | 17 | AQ402095  | AQ402095 HS_5068_A           |
| 35 | 20 | 1.3 | 412 | 17 | BE592272  | BE592272 WS1_50.F0           |
| 36 | 20 | 1.3 | 412 | 17 | A2619913  | A2619913 1M0452G21           |
| 37 | 20 | 1.3 | 422 | 12 | BG044255  | BG044255 sa25h04.            |
| 38 | 20 | 1.3 | 443 | 10 | BE125729  | BE125729 DGI_54.G0           |
| 39 | 20 | 1.3 | 445 | 10 | BE328931  | BE328931 bq32d03.x           |
| 40 | 20 | 1.3 | 455 | 10 | BE329345  | BE329345 bq39g05.x           |
| 41 | 20 | 1.3 | 455 | 10 | BE349696  | BE349696 bq41g04.x           |
| 42 | 20 | 1.3 | 459 | 17 | TA203C07Q | TA203C07Q AL476545 T. brucei |
| 43 | 20 | 1.3 | 460 | 14 | BQ489299  | BQ489299 73-E9180-           |
| 44 | 20 | 1.3 | 475 | 10 | BB772499  | BB772499 BB772499            |
| 45 | 20 | 1.3 | 480 | 13 | BI273111  | BI273111 NF092C06F           |

#### ALIGNMENTS

RESULT 1  
LOCUS CNS00HD3  
DEFINITION Drosophila melanogaster genome survey sequence T83 end of BAC: BACR34P05 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL073332.1 GI:4953112  
VERSION AL073332  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial

#### COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BACP from the isogenic strain y2, can be sp, the same strain used for the BACP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACP/US Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## SOURCE

1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_1lb="BACR34P05"  
/note="end : TERT"

BASE COUNT 327 a 175 c 187 g 333 t 79 others  
ORIGIN

Query Match 1.7%; Score 26; DB 17; Length 1101;  
Best Local Similarity 100.0%; Pred. No. 0.049;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 922 CACAGATTCCTGGATCATATCTT 947  
Db 420 CACAGATTCCTGGATCATATCTT 445

## RESULT 2

## LOCUS

B0118156 529 bp mRNA linear EST 22-JUL-2002  
DEFINITION EST603732 mixed potato tissues Solanum tuberosum cDNA clone STM76

ACCESSION B0118156  
VERSION B0118156.2 GI:21918727

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 529)  
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,  
Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and  
Karanycheva, S.A.  
Generation of a set of potato cDNA clones for microarray analyses  
Unpublished (2002)  
On Apr 17, 2002 this sequence version replaced gi:20170118.  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: [potato@tigr.org](mailto:potato@tigr.org)  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
[cdna@resgen.com](mailto:cdna@resgen.com)  
Seq primer: 73

## FEATURES

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1. 529  
/organism="Solanum tuberosum"  
/cultivar="Kennebec or Binette"  
/db\_xref="taxon:4113"  
/clone\_1lb="STM76"  
/clone\_1lb="mixed potato tissues"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site 1: EcorI, Site 2:  
XhoI; supplier: Combination of untreated and Phytophthora  
infestans-treated libraries of stolons, leaves, leaflets,  
axillary buds of stem explants, petioles, germinating eyes  
tubers, or roots."

## BASE COUNT

## ORIGIN

121 a 137 c 153 g 118 t  
Query Match 1.7%; Score 25; DB 14; Length 529;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 ATGACGACCATGATGATGATGA 92  
Db 470 ATGACGACCATGATGATGATGA 494

## RESULT 3

## LOCUS

A1724181 541 bp mRNA linear EST 19-JUL-2000  
DEFINITION RHIZ1\_8\_B02.y2.A001 Rhizome1 (RHIZ1) Sorghum halepense cDNA, mRNA

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 541)  
Cordonnier-Pratt, M.-M., Gingle, A., Pratt, L.H. and Paterson, A.  
An EST database from Sorghum: Sorghum halepense rhizomes  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: [mpm@plantbio.uga.edu](mailto:mpm@plantbio.uga.edu)  
Sequences have been trimmed to exclude polyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: T7  
High quality sequence stop: 541  
POLY-A-yes.

## FEATURES

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1. 541  
/organism="Sorghum halepense"  
/db\_xref="taxon:4560"  
/clone\_1lb="Rhizome1 (RHIZ1)"  
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda  
Zap II; Site 1: XhoI; Site 2: EcorI; The library was made  
from poly-A RNA in the cloning vector lambda Zap II.  
Clones to be sequenced were prepared by mass excision."  
BASE COUNT 106 a 131 c 180 g 124 t  
ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## OY

## Db

## RESULT 4

## LOCUS

AU267147 302 bp mRNA linear EST 10-MAY-2002  
DEFINITION AU267147 VS Dictyostelium discoideum cDNA clone VSH255 5', mRNA

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

1 (bases 1 to 302)  
Urushihara, H., Morio, T., Saito, T., Koriki, E., Ochiai, H., Maeda, M.,  
Takeuchi, I., Kohara, Y. and Tanaka, Y.  
Population analysis of cDNAs from unicellular and multicellular  
stages of Dictyostelium discoideum



DNA sequence.  
 ACCESSION A2239345  
 VERSION A2239345.1 GI:8552536  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 500)  
 Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akintet, B., Levins, M., McGinn, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C. M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 Other\_GSSs: RPCI-23-32B8.TVB  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@igf.org  
 clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@edje.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (Info@resgen.com). BAC end page: http://www.ligr.org/cdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 32 row: B column: 8  
 Seq primer: SP6  
 Class: BAC ends.  
 FEATURES  
 source Location/Qualifiers  
 1..500  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-32B8"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBAC63.6; Site: 1: ECORI; Site 2: ECORI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of ECORI and ECORI Methylase. Size selected DNA was cloned into the pBAC63.6 vector at the ECORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 152 a 83 c 118 g 147 t  
 ORIGIN  
 Query Match 1.5%; Score 22; DB 17; Length 500;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1292 ATGCTCATGATGCTCATGA 113  
 |||||||  
 Db 479 ATGCTCATGATGCTCATGA 500  
 RESULT 8  
 BJA08775/c 699 bp mRNA linear EST 10-MAR-2002  
 LOCUS BJA08775 Dictyostelium cDNA library, SF Dictyostelium  
 DEFINITION dictostelium cDNA clone dds46m2 3', mRNA sequence.  
 ACCESSION BJA08775  
 VERSION BJA08775.1 GI:19321559  
 KEYWORDS EST  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum.  
 Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 1 (bases 1 to 699)  
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-I, T.  
 Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL Unpublished (2002)  
 COMMENT Contact: Tadasi Shin-I  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 FEATURES  
 source Location/Qualifiers  
 1..699  
 /organism="Dictyostelium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone="dds46m2"  
 /clone\_lib="Dictyostelium discoideum cDNA library, SF"  
 /sex="mat A"  
 /dev\_stage="Slug stage"  
 BASE COUNT 188 a 93 c 154 g 262 t 2 others  
 ORIGIN  
 Query Match 1.5%; Score 22; DB 13; Length 699;  
 Best Local Similarity 100.0%; Pred. No. 5.3;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 77 ATGATGATGATGATGAACAT 98  
 |||||||  
 Db 652 ATGATGATGATGATGAACAT 631  
 RESULT 9  
 BF759175/c 171 bp mRNA linear EST 12-JAN-2001  
 LOCUS BF759175 RC4-CT0477-011200-016-e09 CT0477 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BF759175  
 ACCESSION BF759175.1 GI:12107075  
 VERSION BF759175.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 171)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.E., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 JOURNAL MEDLINE  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-CT0477-011200-016-e09&t3=2000-12-01&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 4  
 High quality sequence stop: 171.  
 Location/Qualifiers  
 1..171  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="CT0477"  
 /dev\_stage="Adult"  
 /note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:

Mon Dec 2 12:19:25 2002

us-09-868-760-6.oli.rst

Page 5

Smal: A mini-library was made by cloning products derived from ONESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 53 a 39 c 42 g 37 t  
ORIGIN

Query Match 1.4%; Score 21; DB 12; Length 171;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 GAGCACCATGATGATGATGAT 90  
Db 126 GAGCACCATGATGATGATGAT 106

RESULT 10  
LOCUS BI702679/c 368 bp mRNA linear EST 18-SEP-2001  
DEFINITION f156h03.y1 zebrafish SJD day 8 fin regeneration Danio rerio cDNA  
ACCESSION BI702679 Clone 4961885 5', mRNA sequence.  
VERSION BI702679.1 GI:15665308  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 368)  
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE Washu zebrafish EST Project 1998  
JOURNAL Unpublished (1998)  
COMMENT Other\_ESTs: f156h03.x2

Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu  
CDNA Library construction by: Joe Barnes and Steve Johnson, DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Research Genetics web address: <http://www.researchgenetics.com/>  
Putative full length read  
The vector to vector length is 369  
zebrafish identity (p-value greater than 1e-99) found to:  
g1124464031gb1AA6057501AA605750 fa18fi2.s1 Ekerrearily gastrulation zebrafish

Seq primer: 73 ET from Amerisham.  
Location/Qualifiers  
1..368  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="4961885"  
/clone\_id="zebrafish SJD day 8 fin regeneration"  
/sex="male"  
/tissue\_type="fin, 8-day regeneration"  
/lab\_host="DH10B"  
/note="Vector: PAMPI. Site\_1: EcoRI; Site\_2: NotI; First strand cDNA synthesis was primed using oligo-dT on magnetic beads with an additional primer  
5'-ggcgccgaatacgaactacata-taggg-3'. Second strand synthesis was a 3-cycle PCR using the primers  
5'-ggcgccgaatacgaactacata-3' and  
5'-aagcgctgtgtacaacgcagagactctt-tttttttttt-3'. CDNA

FEATURES  
source

was subsequently amplified in a 7-cycle PCR with the following primers: 5'-ggcgccgaatacgaactacata-3' and 5'-aagcgctgtgt-tacaacgcag. Deoxy-OMP adaptors were added in a third PCR (5 cycles) and the primers 5'-caucaucaucaagcgccgaatacgaactacata-3' and 5'-caucaucaucaagcgctgtgtgtacgaacgaactac-3'. Ends were treated with uracil DNA glycosylase and product with 3' overhangs was annealed to complementary ends of PAMPI. Insert can be excised using EcoRI and NotI. Library constructed by Joe Barnes and Steve Johnson (Washington University)."

BASE COUNT 122 a 67 c 80 g 99 t  
ORIGIN

Query Match 1.4%; Score 21; DB 13; Length 368;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 CGACCATGATGATGATGATGATCA 92  
Db 32 CGACCATGATGATGATGATGATCA 12

RESULT 11  
LOCUS A0845514/c 387 bp DNA linear GSS 25-MAY-2001  
DEFINITION LMAJFV1.Im26e03.y1 Leishmania major FV1 random genomic library Leishmania major genomic clone LMAJFV1.Im26e03 5' similar to contains element VI-chl\_type\_II.2 leishmania repetitive element ;  
DNA sequence.  
A0845514 GI:6050162  
A0845514 GSS.  
leishmania major.  
leishmania major.  
Eukaryota; Euzenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
1 (bases 1 to 387)

REFERENCE  
AUTHORS Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L., Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A., Blistein, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M., Harvey, N., McCann, R., Tsagaris, H., Williams, T., Jackson, Y., Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.  
A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing: a resource for DNA microarrays and expression profiling  
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)  
21192569  
Other\_GSSs: Im26e03.x1  
Contact: Akopyants, NS / Beverley, SM  
Washu Leishmania Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library construction: Natalia S. Akopyants, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
If using this information please cite:  
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major Friedlin strain VI genome by shotgun sequencing and the Washington University Genome Sequencing Center for information on obtaining clone material please contact: Natalia S. Akopyants Ph.D. (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D. (beverley@borcim.wustl.edu)  
Seq primer: -40RP from gibco  
Class: shotgun  
High quality sequence stop: 370.

JOURNAL MEDLINE  
COMMENT  
TITLE

FEATURES  
source

Location/Qualifiers  
1..387  
/organism="Leishmania major"  
/strain="Friedlin strain VI"

```

/db_xref="taxon:5664"
/clone_1lb="leishmania major Fv1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a light size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
phosphatase and ligated into pZero-2 vector's EcoRV site."
BASE COUNT      42 a      121 c      93 g      129 t
ORIGIN
Query Match      1.4%; Score 21; DB 17; Length 387;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      31 AAGATGACACCATGACGAC 51
        |||||||
Db      299 AAGATGACACCATGACGAC 279

RESULT 12
LOCUS      AM944020/c      461 bp      mRNA      linear      EST 23-APR-2001
DEFINITION      LP03320.Sprime LP Drosophila melanogaster larval-early pupal P0T2
                  Drosophila melanogaster cDNA clone LP03320 3 similar to AC004433;
                  FBPP0002727 sk1l FBgn0016984 FLYBASE:FBPP0002727
                  AC004433:6748 . 9126: 03/15/2001, mRNA sequence.
ACCESSION      AM944020
VERSION
KEYWORDS
SOURCE
ORGANISM      fruit fly.
                  Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 461)
AUTHORS      Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
                  Lewis,S. and Rubin,G.M.
JOURNAL      BDBP/HMT Drosophila EST Project
TITLE      Unpublished (2001)
COMMENT      Other-ESTs: LP03320.Sprime
                  Contact: Stapleton, M.
                  BDBP
                  Lawrence Berkeley National Lab
                  One Cyclotron Rd, Berkeley, CA 94720, USA
                  Fax: 510 486 6798
                  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
                  Based upon the presence of a XhoI site followed by a run of 14 or
                  more T residues at the beginning of the sequence, this clone was
                  polyadenylated. The resulting Poly-T sequence has been removed. hit
                  genomic AE003452: Drosophila melanogaster genomic scaffold
                  14200001386038 section 1 of 15, complete sequence.: 03/15/2001
                  Plate: LP.33 row: B column: 8
                  High quality sequence stop: 367.
FEATURES
source
1..461
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP03320"
/clone_1lb="LP Drosophila melanogaster larval-early pupal
P0T2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: P0T2; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into P0T2. Plasmid cDNA library."
BASE COUNT      82 a      132 c      119 g      128 t
ORIGIN

```

```

Query Match      1.4%; Score 21; DB 10; Length 461;
Best Local Similarity 100.0%; Pred No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      37 GACCACCATGACGACACCAC 57
        |||||||
Db      281 GACCACCATGACGACACCAC 261

RESULT 13
LOCUS      BM159117/c      465 bp      mRNA      linear      EST 03-DEC-2001
DEFINITION      f27d12.y1 zebrafish C32 14 somite embryo Danio rerio cDNA clone
                  5565767 5' similar to SW:YZAL_HUMAN Q16465 HYPOTHETICAL PROTEIN ;,
                  mRNA sequence.
ACCESSION      BM159117.1 GI:17247883
VERSION
KEYWORDS
SOURCE
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
                  ; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 465)
AUTHORS      Clark,M., Johnson,S.L., Lehnach,H., Lee,R., Li,F., Marra,M., Eddy
                  S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
                  K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
                  Waller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E.,
                  Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
                  and Wilson,R.
JOURNAL      WashU zebrafish EST Project 1998
TITLE      Unpublished (1998)
COMMENT      Contact: Stephen L. Johnson
                  Washington University School of Medicine
                  444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                  Tel: 314 286 1800
                  Fax: 314 286 1810
                  Email: zbrafish@wustl.edu
                  cDNA library construction by: Joe Barnes and Steve Johnson. DNA
                  sequencing by: Washington University Genome Sequencing Center Clone
                  distribution: Research Genetics web address:
                  http://www.researchgenetics.com/
                  Possible reversed clone: similarity on wrong strand
                  zebrafish idently (P-value greater than 1e-99) found to:
                  g1125949301gb1AA6587761AA6587761a66a10.s1 zebrafish fin day3
                  regeneration
                  Seq primer: T3 ET from Amerisham.
FEATURES
source
1..465
Location/Qualifiers
/organism="Danio rerio"
/strain="C32"
/db_xref="taxon:7955"
/clone="5565767"
/clone_1lb="zebrafish C32 14 somite embryo"
/tissue_type="embryo, 14 somite"
/note="Vector: pAMP1; Site_1: EcoRI; Site_2: NotI; First
strand cDNA synthesis was primed using oligo-dT on
magnetic beads with an additional primer
5'-ggcgccgtatagctagctacta-3'. Second strand
synthesis was a 3 cycle PCR using the primers
5'-ggcgccgtatagctagctacta-3' and
5'-aagcagtggtacacacgagagctactt-tttttttttt-3'. cDNA
was subsequently amplified in a 7-cycle PCR with the
following primers: 5'-ggcgctaatagctacta-3'
5'-aagcagtggt-aaacacgag. Deoxy-UMP adaptors were added in
a third PCR (5 cycles) and the primers
5'-cauacuucauagcggtacacacgagctactaagg-3' and
5'-cuacuucauagcggtacacacgagagtag-3'. Ends were
treated with uracil DNA glycosylase and product with 3'
overhangs was annealed to complementary ends of pAMP1.
Insert can be excised using EcoRI and NotI. Library
constructed by Joe Barnes and Steve Johnson (Washington

```

```

BASE COUNT      128 a      97 c      104 g      123 t      13 others
ORIGIN

Query Match      1.4%: Score 21; DB 13; Length 465;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 CGACCATGATGATGATGATGA 92
    |||||||
Db 22 CGACCATGATGATGATGATGA 2

RESULT 14
BI946460      493 bp      mRNA      linear      EST 19-OCT-2001
LOCUS      best59.5' Honeybee brain cDNA library Apis mellifera cDNA 5', mRNA
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM      honeybee.
                Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
                Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
                Apoidea; Apis.
                1 (bases 1 to 493)
REFERENCE
AUTHORS      Kucharski, R. and Maleszka, R.
TITLE      Evaluation of differential gene expression during behavioral
JOURNAL      development in the honeybee using microarrays and northern blots
MEDLINE      Genome Biol. 3 (2), research007.1-9 (2002)
COMMENT      Contact: Maleszka R
                Visual Sciences Group
                Research School of Biological Sciences, Australian National
                University
                Biology Pl., Canberra ACT 0200, Australia
                Tel: +61 2 6125 0451
                Fax: +61 2 6125 3784
                Email: maleszka@rsbs.anu.edu.au
                Seq primer: M13 reverse.
                Location/Qualifiers
                1..493
                /organism="Apis mellifera"
                /db_xref="taxon:7460"
                /clone_1db="Honeybee brain cDNA library"
                /sex="Female"
                /tissue_type="central brain, mushroom bodies, central body
                /deuterocerebrum"
                /dev_stage="adult worker"
                /lab_host="E.coli SOLR"
                /note="Organ: brain; Vector: pBluescript II SK-; Site_1:
                EcoRI; Site_2: XhoI; cDNA-harboring phagemids were
                obtained by performing mass excision procedure on honeybee
                brain cDNA lambda unizap XR library (Ebert P.R. et al.,
                1998, Insect Mol. Biol. 7(2):151-62)."
```

```

ACCESSION      AV610738
VERSION
KEYWORDS
SOURCE
ORGANISM      Bos taurus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
                Bovidae; Bovinae; Bos.
                1 (bases 1 to 649)
REFERENCE
AUTHORS      Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H.
TITLE      Establishment of a high throughput EST sequencing system using
JOURNAL      poly(A) tail-removed cDNA libraries and determination of 36,000
MEDLINE      bovine ESTs
COMMENT      Nucleic Acids Res. 29 (22), E108 (2001)
                Contact: Yoshikazu Sugimoto
                Animal Genetics Division
                Shitakawa Institute of Animal Genetics
                Otagawa, Mishige, Nishi-Shitakawa, Fukushima 961-8061, Japan
                Tel: 81-248-25-5641
                Fax: 81-248-25-5723
                Email: kazusugie@occco.ocn.ne.jp
                Single pass sequencing.
                This clone was obtained from a polyA-deleted cDNA library.
                Location/Qualifiers
                1..649
                /organism="Bos taurus"
                /db_xref="taxon:9913"
                /clone_1db="Bos taurus lung fetus"
                /tissue_type="lung"
                /dev_stage="fetus"
                /lab_host="DH10B"
                /note="Vector: pZ1; Site_1: SalI; Site_2: NotI; Poly A
                was deleted from a NotI site"
```

```

BASE COUNT      185 a      118 c      187 g      159 t
ORIGIN

Query Match      1.4%: Score 21; DB 10; Length 649;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 GACCATGATGATGATGATGA 93
    |||||||
Db 591 GACCATGATGATGATGATGA 611
```

Search completed: November 29, 2002, 12:31:13  
Job time : 2078 secs

```

RESULT 15
AV610738      649 bp      mRNA      linear      EST 28-NOV-2001
LOCUS      AV610738 Bos taurus lung fetus Bos taurus cDNA clone E11D034C11 5',
DEFINITION      mRNA sequence.
```

THIS PAGE BLANK (USPTO)